The Effect of Lattice Vibrations on Substitutional Alloy Thermodynamics

by

Axel van de Walle

Submitted to the Department of Materials Science and Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Materials Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2000

Author .....
Department of Materials Science and Engineering
May 8, 2000

Certified by .....
Gerbrand Ceder
Union Minière Associate Professor of Materials Science
Thesis Supervisor

Accepted by .....
Carl V. Thompson II
Stavros Salapatas Professor of Electronic Materials
Chairman, Department Committee on Graduate Students

The Effect of Lattice Vibrations on Substitutional Alloy Thermodynamics

by

Axel van de Walle

Submitted to the Department of Materials Science and Engineering
on May 8, 2000, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Materials Science and Engineering

Abstract

This thesis explores important issues associated with the calculation of phase diagrams from first principles.

A long standing limitation of first-principles phase diagram calculations is the difficulty to account for the impact of lattice vibrations on phase stability. The vast literature addressing this topic is thoroughly reviewed and a clear picture of the origin of vibrational entropy differences between phases in an alloy system is presented. Vibrational entropy change can be attributed to the changes in chemical bond stiffness associated with the changes in bond length that take place during a phase transformation. This so-called "bond stiffness vs. bond length" interpretation both summarizes to key phenomenon driving vibrational entropy changes and provides a practical tool to model them. Accurate first-principles calculations of vibrational entropy differences between ordered and disordered phases are performed in order to confirm the validity of the proposed mechanism.

First-principles calculations of vibrational properties deserve special attention because the widely used Local Density Approximation is known to introduce systematic errors. A semi-empirical correction is introduced and is shown to substantially improve the accuracy of calculated vibrational and elastic properties of extended solids.

This thesis also studies another subproblem associated with the calculation of phase diagrams: the determination of the most stable structures at absolute zero of temperature. Although the technique to find these so-called ground states is well established, its practical implementation remains computationally demanding. A new and useful geometric interpretation to the formal solution of the ground state problem is presented. There is good hope that such intuition will lead to practical algorithm to find the ground states of an alloy system.

Thesis Supervisor: Gerbrand Ceder
Title: Union Minière Associate Professor of Materials Science

Acknowledgments

I would like to thank:

- My advisor, Gerd Ceder, who taught me the art of finding good research topics and writing good papers.
- The members of my thesis committee, Sam Allen and Bob Rose, for their careful reading of my thesis.
- My lab-mates, who kept life interesting by providing numerous occasions to chat about science and life in general.
- My parents, who have always encouraged me to pursue my studies and who were the first to interest me in natural sciences.
- My friends from far away, who did not forget about me for these five years.
- My Fiancée, Susanne, for her support and encouragements, for keeping me from turning into a mad scientist, and for basically everything else ...

À mes parents

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Atomistic simulations

One of the central questions that materials science seeks to answer is: "How to design a material for a specific purpose?" The traditional approach to this question can be summarized as follows:

1. Try out a set of existing materials and pick the best candidates.

2. Fine-tune their properties, by trial and error, through the controlled addition of various impurities or through various processing steps.

A good understanding of the microscopic processes that determine the desired materials properties is of tremendous help in narrowing down the number of possibilities to try out. During the second half of this century, the development of powerful characterization techniques has been a very helpful guide in identifying those mechanisms and in allowing a more effective search for new materials.

Over the last two decades a radically different tool has become available to aid in the materials design: computer simulations at the atomic level. Atomistic simulations offer an unprecedented level of control over experimental conditions and unlimited characterization capabilities. Thanks to this level of control, candidate materials can be tested for desirable properties before considerable effort is spent on attempting to synthesize them. Since the position of each atom and the energy of the system are fully known, any materials property can, in principle, be obtained from an atomistic-level calculation.

In an atomistic simulation, the atoms can, for instance, be placed in a specific configuration so that the formation energy of various point defects can be compared. The knowledge of the precise structure of a defect is crucial to devise ways to limit or promote their formation. Such information is essentially inaccessible experimentally.

There is, however, one main limitation associated with atomistic simulations: Computing power often limits what can be computed at a reasonable cost. For instance, it would be impractical to determine the hardness of a metal by actually simulating, at the atomic level, the whole region of the metal that undergoes plastic

deformation during a conventional hardness measurement. But there is an interesting and important complementarity between virtual and real experiments:

What is difficult to measure is easy to compute ... and vice-versa.

Properties that are difficult to measure include "ideal" properties, such as formation energies of perfect crystals, formation energies of point defects, band structures of perfect crystals, surface energies, activation barriers, etc. These properties are precisely the ones that are easy to calculate, thanks to the high symmetry of the geometries involved. Why focus on these idealized properties? They are the basic ingredients of most theories or models used in materials science. Formation energies, for instance, contribute to determine the relative stability of different phases. Activation barriers determine diffusion coefficients with the help of transition state theory. In the various kinetic models of phase transformations, surface energies influence nucleation and growth rates of precipitates, etc. Knowing these "ideal" properties often brings just the key information that allows a better understanding of the microscopic origin of the macroscopic phenomena of interest.

As computing power increases, increasingly complex properties become accessible to computations, bringing more insight into increasingly intricate phenomena. It is tempting to extrapolate this trend and claim that, one day, any materials properties could be determined by a very accurate brute force calculation involving an unthinkable number of atoms. But this is unlikely to happen, for the following reason. Figuratively speaking, in a real experiment, each nucleus or each electron is responsible for solving the equations governing its behavior. In a numerical experiment, millions of atoms constitute one transistor which can handle only one "bit" of the dozens needed to represent one real number among the hundreds needed to describe the wavefunction of one electron in one of its many possible states. While computer experiments offer better control and characterization capabilities, they are a fundamentally inefficient way to represent exactly what is going on in a real system. The only way that the advantages of computer simulation can be harvested without suffering from their inefficiency is by introducing carefully chosen approximations.

Approximations are possible because most particles in a solid state system spend most of their time doing uninteresting things. Electrons in the lowest energy levels behave nearly as if their were around a free atom. The nuclei oscillate in a nearly harmonic fashion billions of times before exchanging position with a neighboring atom. These uninteresting behaviors can usually be encapsulated in a model that summarizes the essential features. In this fashion, the computational requirements can be drastically reduced and practical results can be obtained. Approximations are a fact of life in calculations, just like measurement errors and impurities are intrinsic to experiments. The main challenge of computational material science is to devise accurate approximations that allow the calculation of materials properties at a reasonable computational cost.

Trying to mimic the actual physical system is almost never the best way to model it.

The present thesis is no exception to that rule in the sense that it seeks develops new approximations that extend the range of properties that can be practically calculated using atomistic calculations.

1.2 Phase diagram calculations

There is currently considerable interest in developing tools that allow the design of materials with the help of virtual experiments. But for such a "virtual laboratory" to be useful, it needs to include a tool that can determine whether a given material can actually be synthesized. This crucial tool is the calculation of phase diagrams. A particular material can only be synthesized if it is stable or, at least, metastable, which is precisely the type of information that is provided by a phase diagram. Phase diagrams have always played a central role in materials science and it is natural that their importance carries over to virtual experiments.

The wealth of thermodynamic information that needs to be calculated in order to determine a phase diagram is also useful in its own respect. Most models determining the kinetic behavior of a system require as an input a "thermodynamic driving force" parameter, which is usually given by a difference in free energy between two phases, an information that is a natural by-product of the calculation of a phase diagram.

In this thesis, we will seek to improve one specific type of phase diagram computations: "First-principles" calculations, also called "*ab initio*" calculations. A first-principles calculation seeks to determine the properties of a material without relying on any experimental input, starting solely from the knowledge of the atomic number of the constituents. We will focus on solid state phase diagrams in alloys, for which first-principles calculations offer the most advantages over experimental measurements. On the one hand, calculations involving solid phases are much more manageable than, for instance, those involving liquid phases. On the other hand, on the experimental side, the determination of the solid state part of a phase diagram is the most likely to be hindered by sluggish kinetics or characterization problems.[1]

The calculation of phase diagrams from first principles is a challenging task, because phase transitions are determined by the collective behavior of a large number of atoms. However, the computational requirements of the quantum mechanical calculations needed to describe the energetics of the system are such that they can only be performed on systems with a small number of atoms.

Over the last twenty years, the formalism allowing the practical calculation of a phase diagram from the information provided by quantum mechanical energy calculations has been carefully laid out. It is now possible to predict relatively complex solid-state phase diagrams from first-principles. Several excellent reviews on the topic exist [41, 39, 150]. However, the accuracy of these calculations is currently limited by a three factors.

[1] While it is easy to determine when an alloy melts, it is much more difficult to determine when it undergoes an order-disorder transition.

1. The quantum mechanical calculations used as an input for the procedure rely on various approximations that limit their accuracy.

2. Computational limitations make it difficult to account for the fact that atoms vibrate around their equilibrium position. For this reason, phase diagram calculations are often performed without including the energetic and entropic effects of atomic vibrations.

3. One of the subproblems associated with the calculation of phase diagrams is the determination of the most stable structures at absolute zero of temperature. The exact solution to this so-called "ground state problem" unfortunately involves substantial computing power.

The problem of the accuracy of quantum mechanical calculations is not unique to phase diagram calculations. Any first-principles calculation of materials properties is potentially limited by this problem. For this reason, this important problem is currently the focus of intense research in the field of solid state physics.

In this thesis, we instead focus on the two last problems, which are more specific to the calculation of phase diagrams. The effect of lattice vibration on alloy thermodynamics is a quite active topic of study within the alloy theory community. The thesis provides a thorough summary of the major findings in this area of research and introduces a new model that embodies the essential physics behind the thermodynamic impact of lattice vibrations. This innovation, called the "bond stiffness vs. bond length" model, not only provides the key intuition behind the origin of vibrational effects, but also suggests a practical way to account for lattice vibrations in phase diagram calculations.

Another original contribution is the introduction of a simple way to alleviate an important problem faced when computing vibrational properties from first-principles: The most widespread approximation used in *ab initio* calculations, called the Local Density Approximation (LDA), results in a systematic bias in the calculated vibrational properties.

This thesis finally addresses the ground state problem by devising a new framework that has the potential to drastically reduce the computational complexity of this task.

1.3 Overview

A review of the current literature focusing on the problem of modeling lattice vibration in first-principles phase diagram calculations is given in Chapter 2. This chapter also introduces the new "bond stiffness vs. bond length" model. Chapter 3 then presents an investigation of two alloy systems that illustrate the effectiveness of the proposed approach, the Ni-Al and Pd-V systems. Chapter 4 describes a simple way to drastically reduce the bias in the calculated vibrational properties caused by the use of the Local Density Approximation. Finally, Chapter 5 attacks the ground state problem and provides a very promising formalism that has the potential of making the determination of the ground states of an alloy a much more manageable problem.

Chapter 2

The Effect of Lattice Vibrations on Phase Stability of Alloys

2.1 Introduction

The field of alloy theory has made substantial progress over the last two decades. It is now possible to predict relatively complex solid-state phase diagrams from first-principles, that is, without the help of any experimental input. Several excellent reviews on the topic exist [41, 39, 150]. The accuracy of calculated phase diagrams is currently limited by two factors. First, one needs, as a starting point, the energy of the alloy in various atomic configurations. The quantum-mechanical calculations used to obtain these energies are computationally intensive and various approximations, such as the local density approximation (LDA), are employed to make calculations tractable. The precision of the calculated phase diagrams is obviously limited by the precision of these approximations.

A second shortcoming is that computational limitations make it difficult to account for the fact that atoms vibrate around their equilibrium position. For this reason, phase diagram calculations are often performed without including the energetic and entropic effects of atomic vibrations. Attempts to either assess the validity of this approximation or to devise computationally efficient ways to account for lattice vibrations are currently the focus of intense research. This interest is fueled by the observation that phase diagrams obtained from first principles often incorrectly predict transition temperatures. It is hoped that lattice vibrations could account for the remaining discrepancies between theoretical calculations and experimental measurements.

Three main questions are addressed in this chapter.

1. Do lattice vibrations have a sufficiently important impact on phase stability that their thermodynamic effects need to be included in phase diagram calculations?

2. What are the fundamental mechanisms that explain the relationship between the structure of a phase and its vibrational properties?

3. How can the effect of lattice vibrations be modeled at a reasonable computational cost?

This chapter is organized as follows. First, Section 2.2 presents the basic formalism that allows the calculation of phase diagrams, along with the generalization needed to account for lattice vibrations. A review of the vast theoretical and experimental literature seeking to quantify the impact of lattice vibrations on phase stability is then presented in Section 2.3. Section 2.4 describes the methods used to calculate vibrational properties while Section 2.5 presents the experimental techniques allowing their measurement. The main mechanisms at the origin of the effect of lattice vibrations on phase stability, presented in Section 2.6, are then used to analyze two common types of approximations: controlled approximations (Section 2.7) and models of lattice vibrations (Section 2.8).

2.2 Generalities

Phase stability at constant temperature is determined by the free energy[1] F. The free energy can be expressed as a sum of a configurational contribution F_{config} and vibrational contributions F_{vib}. The configurational contribution accounts for the fact that atoms can jump from one lattice site to another, while vibrational contribution accounts for the vibrations of each atom around its equilibrium position. The first part of this section presents the traditional formalism used in alloy theory to determine the configurational contribution. The second part introduces the basic quantities that determine whether lattice vibrations have a significant effect on phase stability. The third part describes how the traditional formalism can be adapted when lattice vibrations do need to be accounted for.

2.2.1 Alloy theory

One of the goals of alloy theory is to determine the relative stability of phases characterized by a distinct ordering of atomic species on a given periodic array of sites. This array of sites, called the *parent lattice*, can be any crystallographic lattice augmented by any motif. A convenient representation of an alloy system is the Ising model. In the common case of a binary alloy system, the Ising model consists of assigning a spin-like occupation variable σ_i to each site i of the parent lattice, which takes the value -1 or +1 depending on the type of atom occupying the site. A particular arrangement of spins of the parent lattice is called a *configuration* and can be represented by a vector $\vec{\sigma}$ containing the value of the occupation variable for each site in the parent lattice.

When all the fluctuations in energy are assumed to arise solely from configurational change, the Ising model is a natural way to represent an alloy. The thermodynamics

[1]Strictly speaking, at constant pressure, the Gibbs free energy $G = F + PV$ should be used instead of the Helmoltz free energy F, but at atmospheric pressure, the PV term is negligible for an alloy.

of the system can then be summarized in a partition function of the form:

$$Z = \sum_{\vec{\sigma}} \exp\left(-\beta E(\vec{\sigma})\right) \tag{2.1}$$

where $\beta = 1/\left(k_B T\right)$, and $E(\vec{\sigma})$ is the energy when the alloy has configuration $\vec{\sigma}$. It would be computationally intractable to compute the energy of every configuration from first-principles. Fortunately, the configurational dependence of the energy can be parametrized in a compact form with the help of the so-called cluster expansion [119]. The cluster expansion is a generalization of the well-known Ising Hamiltonian. The energy (per atom) is represented as a polynomial in the occupation variables:

$$\frac{E(\vec{\sigma})}{N} = \sum_{\alpha} m_\alpha J_\alpha \left\langle \prod_{i \in \alpha'} \sigma_i \right\rangle \tag{2.2}$$

where α is a cluster (a set of sites i). The sum is taken over all clusters α that are not equivalent by a symmetry operation of the space group of the parent lattice, while the average is taken over all clusters α' that are equivalent to α by symmetry. The coefficients J_α in this expansion embody the information regarding the energetics of the alloy and are called the effective cluster interaction (ECI). The coefficient m_α are called multiplicities and indicate the number of clusters that are equivalent by symmetry to α (divided by the number of lattice sites).[2] This framework can be extended to arbitrary multicomponent alloys [119].

It can be shown that when *all* clusters α are considered in the sum, the cluster expansion is able to represent any function $E\left(\vec{\sigma}\right)$ of configuration $\vec{\sigma}$ by an appropriate selection of the values of J_α. However, the real advantage of the cluster expansion is that, in practice, it is found to converge rapidly. An accuracy that is sufficient for phase diagram calculations can be achieved by keeping only clusters α that are relatively compact (*e.g.* short-range pairs or small triplets). The cluster expansion thus presents an extremely concise and practical way to model the configurational dependence of an alloy's energy.

The process of calculating the phase diagram of an alloy system can be summarized as follows. First, the energy of the alloy in a relatively small number of configurations is calculated, for instance through first-principles computations. Second, the calculated energies are used to fit the unknown coefficients of the cluster expansion (the ECI J_α). Finally, with the help of this compact representation, the energy of a large number of configurations is sampled, in order to determine the phase boundaries. This can be accomplished with either the Cluster Variation Method (CVM) [69, 41], the low-temperature expansion (LTE) [70], or Monte-Carlo simulations [19].

[2]Both the number of clusters and the number of sites are infinite but their finite ratio can be obtained by ignoring all but one periodic repetitions of the clusters (or the atoms) by the a translational symmetry operation of the lattice.

2.2.2 The effect of lattice vibrations

The previous Section described the framework allowing the calculation of phase diagrams under the assumption that thermodynamic of the alloy is determined solely by configurational excitations. Accounting for vibrational excitations introduces corrections to this simplified treatment. This section presents the basic quantities that enable an estimation of the magnitude of the effect of lattice vibration on alloy thermodynamics. To understand the effect of lattice vibrations on phase stability, it is instructive to decompose the configurational ("config") and vibrational ("vib") parts of the free energy F^α of a phase α into an energetic contribution E and an entropic contribution S:

$$F^\alpha = E^\alpha_{\text{config}} - TS^\alpha_{\text{config}} + E^\alpha_{\text{vib}} - TS^\alpha_{\text{vib}}. \tag{2.3}$$

In the approximation of harmonic lattice vibrations and in the limit of high temperature, the vibrational energy E^α_{vib} is simply determined by the equipartition theorem and is independent of the phase α considered. Hence as long as these approximations are appropriate, lattice vibrations are mainly expected to influence phase stability through their entropic contribution S^α_{vib}.

Intuitively, the vibrational entropy S^α_{vib} is a measure of the average stiffness of an alloy, as can be best illustrated by considering an simple system made of large number of identical harmonic oscillators. The softer the oscillators are, the larger their oscillation amplitude can be, for a fixed average energy per oscillator. Hence, the system samples a larger number of states and the entropy of the system increases. In summary, the softer the alloy, the larger the vibrational entropy.

A phase with a large vibrational entropy is stabilized relative to other phases, since a larger vibrational entropy results in a lower free energy, as seen by Equation (2.3). From a statistical mechanics point of view, this fact can be understood by observing that a phase that encloses more states in phase space is more likely to be visited, as the system undergoes microscopic transitions, and therefore exhibits an increased stability.

The central role of vibrational entropy can be further appreciated by considering the transition temperature between two phases α and β which differ in their vibrational entropy by $\Delta S^{\alpha\to\beta}_{\text{vib}}$. The transition temperature obtained with both configurational and vibrational contributions ($T^{\alpha\to\beta}_{\text{config+vib}}$) is related to the transition temperature obtained with configurational effects only ($T^{\alpha\to\beta}_{\text{config}}$) by

$$T^{\alpha\to\beta}_{\text{config+vib}} \approx T^{\alpha\to\beta}_{\text{config}} \left(1 + \frac{\Delta S^{\alpha\to\beta}_{\text{vib}}}{\Delta S^{\alpha\to\beta}_{\text{config}}}\right)^{-1} \tag{2.4}$$

where $\Delta S^{\alpha\to\beta}_{\text{config}}$ is the change in configurational entropy upon phase transformation [53, 106]. This result[3] indicates that the quantity determining the magnitude of the effect

[3]This result is exact in the limit of small vibrational effects, high temperature and harmonic vibrations.

of lattice vibration on phase stability is the ratio of the vibrational entropy difference to the configurational entropy difference. For this reason, most investigations aimed at assessing the importance of lattice vibrations focus on estimating vibrational entropy differences between phases. Since the configurational entropy (per atom) S_{config} is bracketed by

$$0 \leq S_{\text{config}} \leq -k_B \left(c \ln c + (1-c) \ln (1-c)\right) \leq k_B \ln 2 \approx 0.693 k_B \tag{2.5}$$

for a binary alloy at concentration c, Equations (2.4) and (2.5) provide us with a absolute scale to gauge the importance of vibrations.

2.2.3 Coarse graining of the partition function

The Ising model presented in Section 2.2.1 appears to be intimately associated with the assumption that energy is uniquely determined by the configuration. What can then be done if lattice vibrations do turn out to be important? This section, shows that, in fact, non-configurational sources of energy fluctuations can naturally be taken into account within the Ising model framework through a process called "coarse graining" of the partition function [27, 28].

All the thermodynamic information of a system is contained in its partition function:

$$Z = \sum_i \exp\left[-\beta E_i\right], \tag{2.6}$$

where $\beta = 1/(k_B T)$ and E_i is the energy of the system in state i. In the case of an alloy system, the sum over all possible states of the system can be conveniently factored as following:

$$Z = \sum_L \sum_{\vec{\sigma} \in L} \sum_{v \in \vec{\sigma}} \sum_{e \in v} \exp\left[-\beta E(L, \vec{\sigma}, v, e)\right] \tag{2.7}$$

where

- L is a so-called parent lattice: it is a set of sites where atoms can sit. In principle, the sum would be taken over any Bravais lattice augmented by any motif.
- $\vec{\sigma}$ is a configuration on the parent lattice: It specifies which type of atom rests on each lattice site.
- v denotes the displacement of each atom away from its ideal lattice site.
- e is a particular electronic state (both the spatial wavefunction and spin state) when the nuclei are constrained to be in a state described by v.
- $E(L, \vec{\sigma}, v, e)$ is the energy of the alloy in a state characterized by L, $\vec{\sigma}$, v and e.

Each summation is taken over the states that are included in the set of states defined by the "coarser" levels in the hierarchy of states. For instance, the sum over displacements v includes all displacements such that the atoms remain close to the undistorted configuration $\vec{\sigma}$ on lattice L.

While Equation (2.7) is in principle exact, practical first-principles calculations of phase diagrams typically rely on various simplifying assumptions. The sum over electronic states is often reduced to a single term, namely, the electronic ground state. The validity of this approximation can be assessed by comparing the electronic densities of states at the Fermi level of various structures. If needed, the contribution of electronic entropy is, at least in its one-electron approximation, relatively simple to include without prohibitive computational requirements [145].

An approximation that is much more difficult to justify is the reduction of the sum over displacements v to a single term. This simplification has been extensively used in alloy theory, because calculating the summation over v involves intensive calculations. The particular displacement typically chosen to represent a given configuration $\vec{\sigma}$ is the local minimum in energy that is the closest to the undistorted ideal structure where atoms lie exactly at their ideal lattice sites. In this fashion, the state chosen is the most probable one in the neighborhood of phase space associated with configuration $\vec{\sigma}$. In this approximation, the partition function takes the form of an ising model partition function:

$$Z = \sum_{L} \sum_{\vec{\sigma} \in L} \exp(-\beta E^*(L, \vec{\sigma})) \tag{2.8}$$

with $E^*(L, \vec{\sigma}) = \min_{v,e} \{E(L, \vec{\sigma}, v, e)\}$.

It turns out that the same statistical mechanics techniques developed in the context of the Ising model can also be used in the more general setting where atoms are allowed to vibrate (and where electrons are allowed to be excited above their ground state). All is needed is to replace the energy $E^*(L, \vec{\sigma})$ by the *constrained free energy* $F(L, \vec{\sigma}, T)$, defined as:

$$F(L, \vec{\sigma}, T) = -k_B T \ln \left(\sum_{v \in \vec{\sigma}} \sum_{e \in v} \exp \left[-\beta E(L, \vec{\sigma}, v, e) \right] \right). \tag{2.9}$$

In other words, it is the free energy of the alloy, when its state in phase space is constrained to remain in the neighborhood of the ideal configuration $\vec{\sigma}$. This process, called the "coarse graining" of the partition function, is naturally interpreted as integrating out the "fast" degrees of freedom (*e.g.* vibrations) before considering "slower" ones (*e.g.* configurational changes) [28]. This process is illustrated in Figure 2-1. The quantity to be represented by a cluster expansion is now the constrained free energy $F(L, \vec{\sigma}, T)$. The only minor complication is that the effective cluster interactions become temperature dependent.

There is some level of arbitrariness in the precise definition of the set of displacement v over which the summation is taken in Equation 2.9. However, in the common case where there is a local energy minimum in the neighborhood of $\vec{\sigma}$ and where the

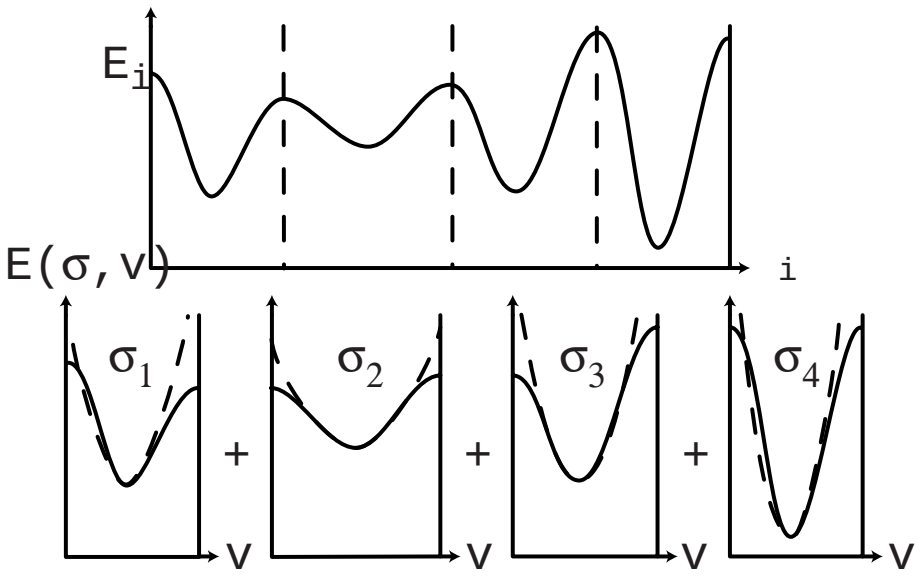


Figure 2-1: The Coarse-Graining Approach.

system spends most of its time visiting a neighborhood that can be approximated by a harmonic potential well, the set of displacements over which the summation is taken has little effect on the calculated thermodynamic properties. Under the above assumptions, calculating the partition function of a constrained harmonic system and a harmonic system of infinite extent gives essentially the same result:

$$\begin{aligned}\sum_{v\in\vec{\sigma}} \exp\left[-\beta E(L,\vec{\sigma},v,T)\right] &\approx \sum_{v\in\vec{\sigma}} \exp\left[-\beta E_H(L,\vec{\sigma},v,T)\right] \\ &\approx \sum_{\text{all } v} \exp\left[-\beta E_H(L,\vec{\sigma},v,T)\right]\end{aligned}$$

where $E(L,\vec{\sigma},v,T) = -k_BT \ln \sum_{e\in v} \exp\left[-\beta E(L,\vec{\sigma},v,e)\right]$ and $E_H(L,\vec{\sigma},v,T)$ denotes a harmonic approximation to $E(L,\vec{\sigma},v)$. In this framework, all is needed to account for lattice vibrations, is the determination of the free energy of a harmonic solid in the neighborhood of any configurations $\vec{\sigma}$. (Appendix A.6 discusses the case where the above assumptions are violated, that is, when no local minimum exists in the phase space neighborhood of $\vec{\sigma}$.)

Although the cluster expansion formalism reduces the intractable problem of determining $F(L,\vec{\sigma},T)$ for *all* $\vec{\sigma}$ to the problem of calculating $F(L,\vec{\sigma},T)$ for a finite set of configurations $\vec{\sigma}$, this task is still much more demanding than calculating the energy $E^*(L,\vec{\sigma})$ for a finite set of $\vec{\sigma}$. Devising an efficient way to calculate $F(L,\vec{\sigma},T)$ is the fundamental problem that needs to be resolved in order to include vibrational effects in phase diagram calculations.

2.2.4 Conclusion

After presenting the basic framework enabling the calculation of the configurational free energy, this section has presented two important aspects of the thermodynamics of lattice vibrations.

1. Vibrational entropy differences between phases introduce corrections to the transition temperatures of an alloy phase diagram calculated including only configurational entropy.

2. The basic alloy theory framework can be adapted to account for lattice vibrations by simply replacing the energy $E^*(L, \vec{\sigma})$ associated with each configuration $\vec{\sigma}$ by the free energy $F(L, \vec{\sigma}, T)$ of a system constrained to remain in the phase space neighborhood of the ideal configuration $\vec{\sigma}$.

2.3 Evidence of vibrational effects

In light of the large computational requirements associated with the inclusion of lattice vibrations, is it important to ensure that such an endeavor is worth the effort. This section reviews the experimental and theoretical evidence that supports the view that vibrational effects are important in the context of phase diagram calculations. There exists a large literature aimed at determining the vibrational properties of solids (see, for instance, [12, 81, 24]). Here, the focus is given to investigations directly related to the determination of vibrational entropy (or free energy) *differences* between phases which differ solely by the ordering of the chemical species on an otherwise *identical parent lattice*.

This relatively narrow choice is driven by two observations. First, while there have been numerous investigations of the absolute vibrational properties of solids, the more difficult issue that needs to be addressed in the context of phase stability is the determination of accurate *differences* in vibrational properties. Second, it has already been established that many structural phase transformations (*e.g.*, from fcc to bcc) are driven by lattice vibrations [111, 59]. This fact does not pose major difficulties for the purpose of phase diagram calculations: one can easily compute the vibrational properties of a few lattice types. The real difficulty is to calculate the vibrational entropy of *many* configurations on *each* of these lattices, a task which is only needed if vibrational properties differ substantially across distinct configurations on an *identical parent lattice*.

The presentation will be mainly chronological, although deviations from that intention will be made for the sake of clarity. We leave a more precise description of the methods used for subsequent sections, focusing here on the results obtained. The key theoretical and experimental results are summarized in Tables 2.1 and 2.3, respectively.

2.3.1 Calculations

The idea that the state of order of an alloy could be coupled with its lattice dynamics is not new. During the 1960's, as the foundations of alloy thermodynamics were being established, the question of the effect of lattice vibrations was already being raised. Studies on the order-disorder transition of β-brass [23, 144], for instance, have indicated that lattice vibrations are crucial to accurately model the magnitude of the experimentally observed discontinuity in heat capacity at the phase transition, which determines the change in vibrational entropy upon disordering.

After these initial investigations, increasingly accurate models for the coupling between lattice vibration and the state of order of an alloy, were then developed [87, 82, 14, 15, 16, 133, 53]. These models generally involved unknown parameters that need to be estimated from available experimental thermodynamic data. A recurring theme among these studies is the idea that, for sensible choices of the stiffness of the springs connecting the atoms, the effect of lattice vibrations is likely to be important. The estimated vibrational entropies of disordering lie between $0.05k_B$, for the most conservative estimates [87], up to the order of $0.5k_B$.[133].

With the increased availability of computing power, the application of first-principles methods became a practical possibility and the unknown parameters of the theoretical models of lattice vibration then became directly computable, without relying on experimental input. Initially, only simple bulk properties, such as the bulk modulus were computable at a reasonable cost. This prompted the development of methods to infer vibrational properties from the knowledge of elastic constants. A particularly popular scheme, the Moruzzi-Janak Schwarz (MJS) method [92], was used in many phase diagram calculations [13, 120, 132, 118, 101, 34, 1, 85, 84, 100, 18]. In the Cd-Mg [13], Ag-Cu [120] and Au-Ni [34] systems, agreement with the experimentally measured phase diagrams was substantially improved.

More recently, other techniques were used to obtain vibrational properties from elastic constants. The so-called virtual crystal approximation (VCA) was used to calculate the vibrational free energy of a disordered alloy in the Ni-Cr system [36]. The calculated vibrational free energy exhibited qualitatively the same concentration-dependence as the vibrational free energy obtained by subtracting the experimentally determined free energy from the calculated configurational free energy.

As computational power steadily increased, the focus of research gradually turned to more accurate treatments of the lattice dynamics. Calculations of the lattice dynamics of ordered and disordered Cu_3Au, within the tight-binding (see, for instance, [62, 112]) and the virtual crystal approximations [32], predicted a vibrational entropy increase of $0.12k_B$ upon disordering. In the first phase diagram calculation based on a full lattice dynamics analysis [129], the SCPIB method [25] was used to calculate that vibrational effects lower the top of the miscibility gap by about 10% in the CaO-MgO system.

In view of the large computational requirements of accurate *ab initio* methods, many researchers sought to calculate vibrational properties with simpler energy models, whose lower computational requirements enabled a more accurate handling of issues such as anharmonicity or the representation of the disordered state. The de-

velopment of the embedded atom method (EAM) [38] offered the opportunity to accurately model metallic alloys at a reasonable computational cost. Investigations based on the EAM typically found large vibrational entropy change upon disordering in metallic alloys. The vibrational entropy change for Cu_3Au was predicted to be $0.10k_B$ [2], while for Ni_3Al, values ranging from $0.22k_B$ to $0.29k_B$ were obtained [2, 8, 9, 115][4].

Other researchers constructed pair potentials from an equation of state determined from first-principles calculations [121]. The resulting pair potentials were then used to calculate disordering vibrational entropies with no further approximations. This method attributed a vibrational entropy change of $0.11k_B$ to the disordering reaction of the Fe_3Al compound.

Rather unexpected results were uncovered as it became possible to compute vibrational entropy differences from a complete lattice dynamics analysis as well as state-of-the-art *ab initio* techniques. Calculations on the Si-Ge system [55] found almost no effect of lattice vibrations: the vibrational entropy of formation of the metastable zincblende structure was a mere $-0.02k_B$. The first *ab initio* calculation of a vibrational entropy of disordering [136] placed an upper bound of $0.05k_B$ in the case of the order-disorder transition of the Ni_3Al compound, in sharp contrast with previous EAM calculations which found a much larger value.

The first phase diagram calculation [106] including *ab initio* vibrational effects, did find a large vibrational effect in the Cu-Au system (about a 20% reduction in transition temperatures). However, the resulting correction on the phase boundaries worsened the agreement with the experimentally determined phase diagram. Interestingly, the same problem had previously been observed in the calculations on the MgO-CaO system [129], although to a lesser extent. Finally, the vibrational entropy change upon disordering of the Pd_3V compound [137] was calculated to be $-0.07k_B$, although the simplest theories put forward in the earliest investigations of lattice vibrations in alloys would predict this value to be large and positive. In short, the latest results based on accurate first-principles calculations represent a significant discontinuity with the earlier evolution of the field. To date, there have been no instances of first-principles phase diagram calculations based on a full lattice dynamics analysis where the effect of vibration was large *and* resulted in an improvement of the agreement with experimental results.

The fact that modeling lattice vibrations does not necessarily improve agreement with experiments points to either one of two possibilities: (i) The discrepancies between calculated and measured phase diagrams arise from effects other than lattice vibrations or (ii) vibrational effects are so difficult to compute accurately that the typical errors made in estimating them are comparable to their absolute magnitude.

[4]Disordered Ni_3Al is actually a metastable phase. The values quoted here are at the highest temperatures reported by the investigators, as close as possible to the true disordering temperature that would be observed if the alloy did not melt before.

Composition	Transition	ΔS $\left(\frac{k_B}{\text{atom}}\right)$	T (K)	Methods	Refs.	year
AgCu	$L1_0^*$ (form)	-0.11	high	ASW,MJS	[120] †	1991
Ag_3Cu	$L1_2^*$ (form)	-0.22	high	ASW,MJS	[120] †	1991
Ag_3Cu	$L1_2^*$ (form)	-0.02	high	ASW,MJS	[120] †	1991
Cu_3Au	$L1_2 \rightarrow$ fcc rnd	0.12	663	TB,BvK,QH,VCA	[32]	1993
CdMg	hcp rnd (form)	0.13	900	LMTO,MJS,CE	[13]	1993
CdMg	B19 (form)	0.14	high	LMTO,MJS	[13] †	1993
Cd_3Mg	$D0_{19}$ (form)	-0.03	high	LMTO,MJS	[13] †	1993
$CdMg_3$	$D0_{19}$ (form)	-0.10	high	LMTO,MJS	[13] †	1993
Cu_3Au	$L1_2 \rightarrow$ fcc rnd*	0.10	high	EAM,BvK,H,SC	[2]	1994
Ni_3Al	$L1_2 \rightarrow$ fcc rnd*	0.29	high	EAM,BvK,H,SC	[2]	1994
SiGe	B3* (form)	-0.02	high	PP,BvK,H	[55]	1996
ArKr	$L1_0$ (form)	-0.06	high	pot.,BvK,H	[55]	1996
$Ca_{0.5}Mg_{0.5}O$	fcc rnd (form)	0.04	high	SCPIB,BvK,H,CE	[129] †	1996
Ni_3Al	$L1_2 \rightarrow$ fcc rnd*	0.27	1400	EAM,BvK,QH,SC	[8][9]	1997
NiCr	fcc rnd (form)	n.a.	1550	FLASTO,VCA	[36]	1997
Ni_3Al	$L1_2 \rightarrow$ fcc rnd*	0.20	1500	EAM,MC,SC	[89]	1998
Ni_3Al	$L1_2 \rightarrow$ fcc rnd*	0.22	1200	EAM,MD,SC	[115]	1998
Ni_3Al	$L1_2 \rightarrow$ fcc rnd*	0.00	high	PP,BvK,QH,SQS	[136]	1998
Ni_3Al	$L1_2 \rightarrow D0_{22}$	0.04	high	PP,BvK,QH	[136]	1998
CuAu	$L1_0 \rightarrow$ fcc rnd	0.18	800	PP,LR,QH,CE	[106]	1998
Cu_3Au	$L1_2 \rightarrow$ fcc rnd	0.08	800	PP,LR,QH,CE	[106]	1998
$CuAu_3$	$L1_2 \rightarrow$ fcc rnd	0.05	800	PP,LR,QH,CE	[106]	1998
CuAu	$L1_1^*$ (form)	0.58	800	PP,LR,QH	[106]	1998
CuAu	$L1_0$ (form)	0.21	800	PP,LR,QH	[106]	1998
Cu_3Au	$L1_2$ (form)	0.20	800	PP,LR,QH	[106]	1998
$CuAu_3$	$L1_2$ (form)	0.26	800	PP,LR,QH	[106]	1998
Al_3Li	$L1_2^* \rightarrow D0_{22}^*$	0.04	high	PP,BvK,H	[123]	1999
Fe_3Al	$D0_3 \rightarrow$ bcc rnd	0.11	high	TB-LMTO,pot.,H,SC	[121]	1998
Pd_3V	$D0_{22} \rightarrow$ fcc rnd	-0.07	high	PP,BvK,QH,SQS	[137]	2000
Pd_3V	$L1_2 \rightarrow D0_{22}$	0.08	high	PP,BvK,QH	[137]	2000

(form): Vibrational entropy of formation from pure elements.
rnd: Disordered solid solution.
$*$: metastable compound
†: calculated from the data presented in the paper.

Table 2.1: Calculated Vibrational Entropy Differences. See Table 2.2 for abbreviations.

pot.: pair potentials	H: harmonic approximation
EAM: embedded atom method	QH: quasiharmonic approximation
TB: tight-binding	MC: Monte Carlo
TB-LMTO: tight-binding LMTO	MD: molecular dynamics
LMTO: linear muffin-tin orbitals	D: Debye model
ASW: atomic spherical waves	BvK: Born-von Kármán spring model
PP: pseudopotential calculations	SC: supercell method
	LR: linear response
cal.: differential calorimetry measurements	VCA: virtual crystal approximation
1xtal: single crystal phonon dispersion measurements	MJS: Moruzzi, Janak & Schwarz model
INS: incoherent neutron scattering measurements	CE: cluster expansion
anh.: anharmonicity included.	SQS: special quasirandom structures

Table 2.2: Abbreviations Used in Tables 2.1 and 2.3.

2.3.2 Comparison with Experiments

Over the last 10 years, advances in experimental techniques have made it possible to directly measure vibrational entropy differences, instead of inferring them from discrepancies between measured thermodynamical data and calculated estimates of configurational contributions to the free energy. Experimental investigations have thus provided independent assessments of the role of lattice vibration on phase stability.

The first direct measurement was obtained from differential calorimetry measurements on the Ni_3Al compound [11] and found the vibrational entropy of disordering to be at least $0.19k_B$. This finding was corroborated with subsequent incoherent neutron scattering measurements [49], which bracketed its value between $0.1k_B$ and $0.3k_B$. These findings fueled much of the interest of the recent theoretical literature on the Ni_3Al compound [2, 8, 9, 115, 136]. Unfortunately, the agreement between experimental and theoretical determinations is relatively poor: Even in studies where the vibrational entropy itself agrees quantitatively, its proposed physical origin differs substantially: Experiments explained the vibrational entropy change within the harmonic approximation, while most calculations [8, 9, 115] attributed it to anharmonic effects. Many of these conflicting findings were clarified by first-principles calculations [136], which identified the problems encountered in the experimental measurements and the earlier calculations. The general consensus among researchers is now that given the numerous difficulties faced when studying Ni_3Al, unambiguous evidence of the importance of lattice vibrations should probably be sought in other systems.

Similar calorimetry measurements were then performed on the Fe_3Al [10] and Cu_3Au [95] compounds, where more conclusive results could be obtained. The vibrational entropy of disordering of Fe_3Al was determined to be $0.10k_B$, a result which was later corroborated by calculations [121]. In the case of Cu_3Al the experimental result, $0.14k_B$, showed very good agreement with the earlier theoretical predictions of $0.12k_B$ [32].

The estimation of vibrational effects was also addressed by directly probing the lattice dynamics through neutron scattering measurements. The vibrational entropy

of formation of a disordered alloy in the Fe-Cr system was obtained from single crystal measurements of phonon dispersion curves [50] in the virtual crystal approximation.Values ranging from $0.14k_B$ to $0.21k_B$ were obtained, depending on concentration.

In order to determine the lattice dynamics of disordered alloys beyond the virtual crystal approximation, the incoherent neutron scattering technique was extensively used and refined [98, 97, 117, 21, 49]. With this technique, the analysis of the experimental data is considerably simplified when the species present have comparable incoherent neutron scattering intensities, which lead to the study of two compounds satisfying this requirement: Ni_3V and Co_3V. Measurements on the Ni_3V compound [98] found a surprisingly small vibrational entropy change upon disordering, $0.04k_B$, while the Co_3V compounds exhibited a relatively large value $0.15k_B$ [97]. A related study found the vibrational entropy change associated with the fcc-hcp transition of the Co_3V compound to be $0.07k_B$ [117]. It is interesting to note that the disordering reaction exhibits a larger vibrational entropy change than the allotropic transformation in Co_3V. Perhaps more importantly, the investigations of the Ni_3V and Co_3V compounds presented the first experimental evidence of important anharmonic effects.

The same technique of incoherent neutron scattering was employed to revisit the Cu-Au system [21]. The vibrational entropy of formation of the Cu_3Au compound was found to be $0.06k_B$ at $300K$, corroborating earlier estimations based on phonon dispersion curve measurements [20]. At $800K$, the measured value of $0.12k_B$ falls within experimental error of the results of ab-initio calculations of $0.20k_B$ [106].

The vibrational entropy of formation of various ordered compounds, obtained from single crystal phonon dispersion measurements, were recently compiled [20] and show formation values of up to $0.5k_B$. However, this compilation contains many systems where the alloy has a crystal structure that differs from the one of the pure elements and the formation values thus also include the vibrational entropy change associated with a structural transition. When all these cases are excluded, the maximum vibrational entropy change decreases to a more conservative upper bound of $0.20k_B$, which is reached in the ordered phase of Ni_3Al.

2.3.3 Conclusion

Although early investigations of the impact of vibrational effects on phase stability consistently found large effects, it is now becoming apparent, as more precise theoretical and experimental techniques became available, that vibrational effects are not systematically large. It is therefore important to identify the factors which determine when they are, so that the effort devoted to calculating them is proportional to their expected magnitude.

2.4 Computational techniques

To understand the nature of the difficulties encountered, it is instructive to first consider how, in principle, the vibrational properties of a single configuration $\vec{\sigma}$ can be calculated with an arbitrary accuracy. The techniques presented in this section are

Composition	Transition	ΔS $\left(\frac{k_B}{\text{atom}}\right)$	T (K)	Methods	Refs.	year
Ni_3Al	$L1_2 \rightarrow$ fcc rnd	0.27	high	cal.,D,H	[11]	1993
Ni_3Al	$L1_2 \rightarrow$ fcc rnd	0.19	343	cal.	[11]	1993
Fe_3Al	$D0_3 \rightarrow$ bcc rnd	0.10 ± 0.03	high	cal.,D,H	[10]	1994
Cu_3Al	$L1_2 \rightarrow$ fcc rnd	0.14 ± 0.05	high	cal.,1xtal,H,VCA	[95]	1995
$Fe_{0.70}Cr_{0.30}$	bcc rnd (form)	0.14 ± 0.05	high	1xtal,H,VCA	[50]	1995
$Fe_{0.53}Cr_{0.47}$	bcc rnd (form)	0.20 ± 0.05	high	1xtal,H,VCA	[50]	1995
$Fe_{0.30}Cr_{0.70}$	bcc rnd (form)	0.21 ± 0.05	high	1xtal,H,VCA	[50]	1995
Ni_3Al	$L1_2^* \rightarrow$ fcc rnd	0.10	high	INS,H	[49]	1995
Ni_3Al	$L1_2^* \rightarrow$ fcc rnd	0.30	high	INS,H,VCA	[49]	1995
Ni_3V	$D0_{22} \rightarrow$ fcc rnd	0.04 ± 0.02	300	cal.,INS	[98]	1996
Co_3V	$L1_2^* \rightarrow$ fcc rnd	0.15 ± 0.02	high	INS	[97]	1997
Cu_3Au	$L1_2$ (form)	0.06 ± 0.03	300	INS,anh.	[21]	1999
Cu_3Au	$L1_2$ (form)	0.12 ± 0.03	800	INS,anh.	[21]	1999
Co_3V	hP24 $\rightarrow$ fcc rnd	0.07	high	INS,QH	[117]	1999
$CeSn_3$	γ-Ce + β-Sn $\rightarrow L1_2$	-0.54 ± 0.09	high	1xtal,H	[20]	1999
$LaSn_3$	hcp-La + β-Sn $\rightarrow L1_2$	-0.43 ± 0.09	high	1xtal,H	[20]	1999
Ni_3Al	$L1_2$ (form)	-0.20 ± 0.03	high	1xtal,H	[20]	1999
Ni_3Fe	fcc-Ni + bcc-Fe $\rightarrow L1_2$	0.09 ± 0.03	high	1xtal,H	[20]	1999
Pt_3Fe	fcc-Pt + bcc-Fe $\rightarrow L1_2$	0.14 ± 0.03	high	1xtal,H	[20]	1999
Pd_3Fe	fcc-Pd + bcc-Fe $\rightarrow L1_2$	0.05 ± 0.03	high	1xtal,H	[20]	1999
Cu_3Zn	fcc-Cu + hcp-Zn $\rightarrow L1_2$	-0.01 ± 0.03	high	1xtal,H	[20]	1999
Cu_3Au	$L1_2$ (form)	0.07 ± 0.03	high	1xtal,H	[20]	1999
Fe_3Pt	bcc-Fe + fcc-Pt $\rightarrow L1_2$	0.55 ± 0.03	high	1xtal,H	[20]	1999
Fe_3Al	bcc-Fe + fcc-Al $\rightarrow D0_3$	-0.06 ± 0.03	high	1xtal,H	[20]	1999

$*$: metastable compound

Table 2.3: Experimental Measurements of Vibrational Entropy Differences. See Table 2.2 for abbreviations.

the tools that were used to investigate the importance of lattice vibrations presented in Section 2.3.1.

Phase diagram calculation involves computing vibrational properties for a set of configurations $\vec{\sigma}$. Carrying out the full phonon problem for each configuration results in undue computational requirements. Nevertheless the formal solutions presented here play an important in devising practical ways to include vibrational effects in phase diagram calculations. This section first focuses on the treatment lattice vibration within the harmonic approximation, before addressing the issue of anharmonicity. Finally, important consideration concerning the energy models used as an input for these procedures are discussed.

2.4.1 Lattice vibrations in the harmonic approximation

In this section, we review the problem of determining the constrained free energy of a system in the neighborhood of a configuration $\vec{\sigma}$, under the assumption that the system spends most of its time in a region near a local energy minimum, where a harmonic approximation to the energy surface is accurate. In this approximation, the free energy determination reduces to the well-known phonon problem [81], [12].

Theory

Consider a system consisting of N atoms. Let M_i be the mass of atom i and $u(i)$ be its displacement away from its equilibrium positions. Time derivatives are denoted by dots while Greek letter subscripts denote one of the cartesian components of a vector. In the harmonic approximation, the energy of the system can be written as:

$$H = \frac{1}{2}\sum_i M_i \left(\dot{u}(i)\right)^2 + \frac{1}{2}\sum_{i,j} u^T(i)\,\Phi(i,j)\,u(j)$$

where

$$\Phi_{\alpha\beta}(i,j) = \left.\frac{\partial^2 E}{\partial u_\alpha(i)\,\partial u_\beta(j)}\right|_{u(l)=0\ \forall l}.$$

The 3×3 matrices $\Phi(i,j)$ are called the force constants tensors, as they relate the displacement of atom j to the force f exerted on atom i:

$$f(i) = \Phi(i,j)\,u(j).$$

Such a harmonic approximation of a solid is often referred to as a Born-von Kármán model.

Note that contrary to usual treatment, we do not immediately impose translational symmetry, in order to derive a few general results that also apply to systems such as disordered alloys.

The substitution $e(i) = \sqrt{M_i}u(i)$ yields:

$$H = \frac{1}{2}\left(\sum_i \dot{e}^2(i) + \sum_{i,j} e^T(i)\frac{\Phi(i,j)}{\sqrt{M_i M_j}}e(j)\right). \tag{2.10}$$

The $3N$ eigenvalues λ_m of the matrix[5]

$$D = \begin{pmatrix} \frac{\Phi(1,1)}{\sqrt{M_1 M_1}} & \cdots & \frac{\Phi(1,N)}{\sqrt{M_1 M_N}} \\ \vdots & \ddots & \vdots \\ \frac{\Phi(N,1)}{\sqrt{M_N M_1}} & \cdots & \frac{\Phi(N,N)}{\sqrt{M_N M_N}} \end{pmatrix}. \tag{2.11}$$

then give the frequencies $\nu_m = \frac{1}{2\pi}\sqrt{\lambda_m}$ of the normal modes of oscillation. In the harmonic approximation, the knowledge of these frequencies is sufficient to determine the thermodynamic quantities we are interested in. This information is conveniently summarized by the phonon density of states (DOS), which gives the number modes of oscillation having a frequency lying in the interval $[\nu, \nu + d\nu]$:

$$g(\nu) = \frac{1}{N}\sum_{m=1}^{3N}\delta(\nu - \nu_m).$$

It can be shown that the free energy of the system (restricted to remain close to a given configuration $\vec{\sigma}$) is given by [81]:

$$\begin{aligned} \frac{F}{N} &= \frac{E^*}{N} + k_B T\int_0^\infty \ln\left(2\sinh\left(\frac{h\nu}{2k_B T}\right)\right) g(\nu)\, d\nu \\ &= \frac{E^*}{N} + \frac{k_B T}{N}\sum_m \ln\left(2\sinh\left(\frac{h\nu_m}{2k_B T}\right)\right) \end{aligned}$$

where E^* is the potential energy of the system at its equilibrium position and h is Planck's constant. Phase transitions in alloys typically occur at a temperature where $\frac{h\nu_m}{k_B T} \ll 1$ for most ν_m, so that the high temperature limit of this expression is a useful simplification:

$$\begin{aligned} \frac{F}{N} &= \frac{E^*}{N} + k_B T\int_0^\infty \ln\left(\frac{h\nu}{k_B T}\right) g(\nu)\, d\nu \\ &= \frac{E^*}{N} + \frac{k_B T}{N}\sum_m \ln\left(\frac{h\nu_m}{k_B T}\right) \end{aligned}$$

The usual criterion used to determine the temperature range where high temperature

[5]Among the $3N$ eigenvalues, the 6 eigenvalues associated with rigid body translations and rotations are zero. In the thermodynamic limit, these few degrees of freedom are inconsequential. To avoid notational complications, we simply assume that the solid is fixed to a reference frame by springs so that the resulting dynamical matrix has $3N$ nonzero eigenvalues.

limit is reached is the Debye temperature. Note that the factor $\frac{h}{k_B T}$ is often omitted because it cancels out when calculating vibrational free energy differences. In the high temperature limit, another important form of cancellation occurs: The atomic masses have no effect on the free energies of formation. This important result, shown in Appendix A.1, rules out that masses play any significant role in determining phase stability at high temperatures.

As mentioned before, a convenient measure of the magnitude of the effect of lattice vibrations on phase stability is the vibrational entropy, which can be obtained from the vibrational free energy by the well known thermodynamical relationship $S_{\text{vib}} = -\frac{\partial F_{\text{vib}}}{\partial T}$. Contrary to the vibrational free energy of formation, the vibrational entropy of formation[6] is temperature-independent in the high-temperature limit of the harmonic approximation, allowing a unique number to be reported as a measure of the importance of vibrational effects.

In a crystal, the determination of the normal modes is somewhat simplified by the translational symmetry of the system. Let n denote the number of atoms per unit cell. Let $u\binom{l}{i}$ denote the displacements away from its equilibrium position of atom i in cell l. Let $\Phi\binom{l\ l'}{i\ j}$ be the force constant relative to atom i in cell l and atom j in cell l' and let $e\binom{l}{i} = \sqrt{M_i} u\binom{l}{i}$. Bloch's theorem indicates that the eigenvectors of the dynamical matrix are of the form

$$e\binom{l}{i} = e^{\iota 2\pi(k\cdot l)} e\binom{0}{i} \tag{2.12}$$

where l denotes the cartesian coordinates of one corner of cell l and k is a point in the first Brillouin zone. This fact reduces the problem of diagonalizing the $3N \times 3N$ matrix D to the problem of diagonalizing a $3n \times 3n$ matrix $D(k)$ for various values of k. This can be shown by a simple substitution of Equation (2.12) into Equation (2.10).[7] The dynamical matrix $D(k)$ to be diagonalized is given by[8]

$$D(k) = \sum_l e^{i2\pi(k\cdot l)} \begin{pmatrix} \frac{\Phi\binom{0\,l}{1\,1}}{\sqrt{M_1 M_1}} & \cdots & \frac{\Phi\binom{0\,l}{1\,n}}{\sqrt{M_1 M_n}} \\ \vdots & \ddots & \vdots \\ \frac{\Phi\binom{0\,l}{n\,1}}{\sqrt{M_n M_1}} & \cdots & \frac{\Phi\binom{0\,l}{n\,n}}{\sqrt{M_n M_n}} \end{pmatrix}.$$

As before, the resulting eigenvalues $\lambda_i(k)$ for $i = 1 \ldots n$, give the frequencies of the normal modes ($\nu_i(k) = \frac{1}{2\pi}\sqrt{\lambda_i(k)}$). The function $\nu_i(k)$ for a given i is called a phonon branch, while the plot of the k-dependence of all branches along a given direction in k space is called the phonon dispersion curve. In periodic systems, the phonon DOS

[6]The absolute value of the vibrational entropy is not constant at high temperature, but its temperature-dependence does not vary across distinct phases and thus formation values are temperature-independent.

[7]And changing the summation over atoms by summations over atoms and cells.

[8]The reader shown be aware that there are many possible conventions regarding the phase factor: for instance, $e^{i2\pi(k\cdot l)}$, $e^{-i2\pi(k\cdot l)}$, $e^{i(k\cdot l)}$, $e^{i2\pi(k\cdot(l+x(j)))}$ where $x(j)$ is the coordinate of atom j within the cell. While all convention yield different dynamical matrices, they all have the same eigenvalues.

is defined as

$$g(\nu) = \sum_{i=1}^{3n} \int_{BZ} \delta(\nu - \nu_i(k))\, dk$$

where the integral is taken over the first Brillouin zone.

Force Constant Determination

The above theory relies, of course, on the availability of the force constant tensors. The determination of these force constant tensors is the focus of this section. Before describing the methods used for their determination, we will first review important properties of the force constant tensors.

While the number of unknown force constants to be determined is in principle infinite, it can, in practice, be reduced to a manageable finite number with the help of the following two observations. First, the force constant $\Phi(i,j)$ between two atoms i and j beyond a given distance can be neglected. Second, the symmetry of the crystal imposes linear constraints between the elements of the force constant tensors.

The accuracy of the approximation made by truncating the range of force constant can be tested by gradually increasing the range of interactions, until the quantities to be determined no longer vary substantially. It is important to note that most thermodynamic quantities can be written as a weighted integral of the phonon DOS and their convergence rates are thus much faster than the pointwise convergence rate of the phonon DOS itself [55, 136]. That is, the errors on the DOS at each frequency tend to be quickly averaged out when the contributions of each frequency are added.

The restrictions on the force constants imposed by the symmetry of the lattice can be expressed as follows. Consider the force constant $\Phi(i,j)$ of atoms i and j located at $x(i)$ and $x(j)$ and consider a symmetry transformation that maps a point of coordinate x to $Sx+t$, where S is a 3×3 matrix and t and 3×1 translation vector. In general, if the crystal is left unchanged by such a symmetry operation, the force constant tensors should be left unchanged as well. This fact imposes the following constraints on the spring tensors:

$$\left.\begin{array}{l} Sx(i)+t = x(i') \\ Sx(j)+t = x(j') \end{array}\right\} \Rightarrow \Phi(i,j) = S^T \Phi(i',j') S.$$

Additional constraints on the force constants can be derived from simple invariance arguments. The most important constraints, obtained by noting that rigid translations and rotations must leave the forces exerted on the atoms unchanged, are

$$\begin{aligned} \Phi(i,i) &= -\sum_{j\neq i} \Phi(i,j) \\ \Phi(i,j) &= \Phi^T(j,i) \end{aligned}$$

Additional constrains can be found in [81, 24].

There are essentially three approaches to determining the force constants: analytic calculations, supercell calculations and linear response calculations. Analytic calculations are only possible when the energy model is sufficiently simple to allow a direct calculation of the second derivatives of the energy with respect to atomic displacements, as in the case of empirical pair potential models. For first-principles calculations, either one of the two following methods have to be used.

The supercell method The supercell method [143],[55] consists of slightly perturbing the positions of the atoms away from their equilibrium position and calculating the reaction forces. Equating the calculated forces to the forces predicted from the harmonic model yields a set of linear constraints that allows the unknown force constants to be determined.[9]

$$F(i) = \Phi(i,j)u(j)$$

When the force constants considered have a range that exceeds the extent of the primitive cell, a supercell of the primitive cell has to be used. (The simultaneous movement of the image atoms introduces linear constrains among the forces that prevent the determination of some of the force constants.)

While any choice of the perturbations that allows the force constants to be determined is in principle equally valid, a few simple principles drastically narrow down the number of perturbations that need to be considered. For a given supercell, there is a only of finite number of non-redundant perturbations to consider.

A minimal set of non-redundant perturbations can be obtained as follows.

- Consider in turn each atom in the asymmetric unit of the primitive cell.
- Mark the chosen atom (and its periodic images in the other supercells) and consider it as distinct from other atoms of the same type. (This operation effectively removes some of the symmetry operation of the space group of the crystal.)
- Construct the point group $\{S_i\}$ of the site where this atom is located. (S_i is a 3×3 matrix.)
- Move the chosen atom along a direction u_1 such that the space spanned by the vectors $S_i u_1$ (for all i) has the highest dimensionality possible.
- If the resulting dimensionality is less than three, consider an additional direction u_2 such that the space spanned by the vectors $S_i u_j$ for $j = 1, 2$ has the highest dimensionality possible.

[9]Equalities between calculated and predicted energies can be used as well. Using energies alone to determine the force constants would be a rather inefficient use of the information provided by ab-initio calculations. Once a first-principles calculations of the energy of a distorted structure has been completed, the calculation of the forces acting on the atoms is computationally inexpensive. The knowledge of the energy provides a single equation while the knowledge of the forces provide up to 3 equations per atom.

- If the resulting dimensionality is less than three, consider a direction u_3 orthogonal to u_1 and u_2.

The resulting displacements u_j for all atoms in the asymmetric unit gives a minimal list of perturbations that is sufficient to find all the force constants that can possibly be determined with the given supercell. This result follows from the observation that any other possible displacement can be written as a linear combination of the displacements considered above (or displacements that are symmetrically equivalent to them).

When determining force constants with the supercell method, it is important to verify that the presence of small numerical noise in the calculated forces does not result in too much error in the fitted force constants. To minimize noise in the fitted force constants, it may be necessary to use more than the minimum possible number of perturbations. The additional perturbations should ideally be based on different supercells, to minimize the systematic errors introduced by the movement of the image atoms.

When ab-initio calculations are used to calculate the forces, it is especially important to iterate the electronic self-consistency steps to convergence. Even though the energy may appear to be well converged, the forces may not yet be. Energy is the solution to a minimization procedure, while forces are not. As a result, errors on the energy are of a second order in the minimization parameters, while the errors on the forces are of the first order in the minimization parameters. For the same reason, special attention should be given to the structural relaxations.

The true system is not exactly harmonic and the calculated forces may exhibit anharmonic components that introduce noise into the fitted force constants. This problem can be alleviated by considering an additional set of perturbations, where the displacements have the opposite sign. Subtracting the calculated forces obtained for this new set of displacements from the corresponding displacements of the opposite sign exactly cancels out all the odd-order anharmonic terms. Of course, for perturbations such that the negative displacement is equivalent by symmetry to the corresponding positive displacement, this duplication is unnecessary, because the terms of odd order are already zero by symmetry.

Additional guidelines for fitting force constants can be found in [3, 143, 55].

Linear response Linear response calculations seek to directly evaluate the dynamical matrix for a set of k points. The starting point of the linear response approach is evaluation of the second-order change in the electronic energy induced by a atomic displacements from perturbation theory. Within this framework, practical schemes to compute vibrational properties in semiconductors [17, 56, 57, 142] and metallic systems [40, 114, 104] have been devised. In this section we will not discuss the theory behind linear response calculations, but rather focus on how the results of linear response calculations can be used in the context of alloy phase diagram calculations.

The dynamical matrices calculated from linear response theory are exact in the sense that they account for arbitrarily long-range force constants. While in the supercell method inaccuracies arise from the truncation of the force constants, the limit

in precision for linear response calculations arises from the use of a small number of k points to sample the Brillouin zone. To address this issue, two methods can be used.

A set of special k points can be selected through the Chadi-Cohen [30] or Monkhorst-Pack [86] schemes. Special k points are selected so that the integral over the Brillouin zone of a function $f(k)$ that contains no Fourier components above a given frequency can be exactly evaluated by a weighted average of the function at each special point. Since thermodynamic quantities can be written as integrals of functions of the dynamical matrix $f(D(k))$ over the Brillouin zone, the procedure is straightforward to apply in this context.

The other approach is the so-called Fourier inversion method (see, for instance, [56, 114]). The calculated dynamical matrices from a set of k points are used to determine the value of the force constants up to a certain interaction range. The resulting harmonic model can then be used to calculate the dynamical matrix at any point in the Brillouin zone, allowing a much finer sampling of the Brillouin zone for the purpose of performing the numerical integration required to determine any thermodynamic quantity.

The Fourier inversion method is preferable when the function $f(D(k))$ to be integrated exhibits high-frequency components, while the dynamical matrix itself, $D(k)$, does not. Such a situation would arise when $f(\cdot)$ is highly nonlinear. The smoothness of $D(k)$ then ensures that it can be represented with a small number of Fourier components. The less well-behaved function $f(D(k))$ can then be accurately integrated with as many k points as needed, using the dynamical matrix $D(k)$ calculated from the spring model.

In the case of vibrational free energy calculations, the special k points method has been observed to converge rapidly with respect to the number of k points [54][10], so that the Fourier inversion method is probably unnecessary.[11]

For a given set of special k points, there is an approximate correspondence between the number of Fourier components that can be integrated exactly and the range of force constants that can be determined. The correspondence is exact only when the lattice has one atom per cell and when the function $f(\cdot)$ is linear.[12]

While supercell and linear response calculations are in principle equivalent in terms of the information they provide, they have complementary advantages in terms of computational efficiency. The linear response method is the most efficient way to perform high-accuracy calculations that would otherwise be tedious and computer intensive with the supercell method. However, when a high accuracy is not needed, the supercell method has the advantage that various simplifying assumptions regarding the structure of the force constant tensors can transparently be used to drastically

[10] Calculations of the authors, based on data from [106] also support this finding.

[11] Note that the function whose integral gives the free energy exhibits a logarithmic singularity at the Γ point, which could lead to high frequency components that are difficult to integrate accurately. However, in three-dimensional systems, this logarithmic singularity contributes very little to the value of the free energy, so that the rate of convergence of the integral as a function of the number of k points is not dramatically slowed down by the presence of the singularity. As a result, the special k point method can safely be used in practical calculations of the vibrational free energy.

[12] As can be shown by a simple Fourier transform.

reduce computational requirements. It is not clear at which level of accuracy the cross-over between the efficiency of each approach occurs, but it is important to keep both approaches in mind. Another consideration is that in the continuously evolving field of computational solid state physics, new first-principles energy methods are continually developed, and the derivation of the appropriate linear response theory always follows the derivation of simple force calculations. Hence, despite the elegance of linear response theory, it is to be expected that the supercell method will always remain of interest.

2.4.2 Anharmonicity

While the harmonic approximation is remarkably accurate given its simplicity, it has one important limitation: It is unable to model thermal expansion and its impact on vibrational properties. Both the free energy F and the entropy S can be obtained from the the heat capacity C_p:

$$S = \int_0^T \frac{C_p}{T} dT \text{ and } F = E|_{T=0} - \int_0^T S\, dT,$$

Hence, a simple way to account for thermal expansion is to use the following well known thermodynamic relationship between the heat capacity at constant pressure C_p and at constant volume C_v:

$$C_p = C_v + BVT\alpha^2 \tag{2.13}$$

where α is the coefficient of volumetric thermal expansion while B is the bulk modulus. In a purely harmonic model, there is no thermal expansion and C_v can be regarded as the heat capacity of a harmonic model while the correction term $BVT\alpha^2$ arises from anharmonic effects.

Equation (2.13) is directly useful in the context of experimental measurements where all quantities can be directly measured [98, 117]. In the following section, we describe the computational techniques used to handle anharmonicity.

The quasiharmonic model

A simple modification to the harmonic approximation, called the quasiharmonic approximation, allows the calculation of thermal expansion at the expense of a moderate increase in computational cost. In the quasi-harmonic approximation, the phonon frequencies are allowed to be volume-dependent, which amounts to assuming that the force constant tensors are volume-dependent. The best way to understand this approximation is to study a simple model system where it is essentially exact. Consider a linear chain (with periodic boundary conditions) of identical atoms interacting solely with their nearest neighbors through a pair potential of the form:

$$U(r) = a_1 r + a_2 r^2 + a_3 r^3.$$

Let L be the average distance between two nearest neighbors and let $u(i)$ denote the displacement of atom i away from its equilibrium position. The total potential energy (per atom) of this system is then given by

$$\begin{aligned}\frac{U}{N} &= \frac{1}{N}\sum_i a_1\left(L+u(i)-u(i+1)\right)+a_2\left(L+u(i)-u(i+1)\right)^2 \\ &\quad +a_3\left(L+u(i)-u(i+1)\right)^3\end{aligned}$$

This expression can be simplified by noting that all the terms that are linear in $(u(i)-u(i+1))$ cancel out when summed over i.

$$\begin{aligned}\frac{U}{N} &= a_1L+a_2L^2+a_3L^3+(a_2+3a_3L)\frac{1}{N}\sum_i\left(u(i)-u(i+1)\right)^2 \\ &\quad +O\left(\left(u(i)-u(i+1)\right)^3\right)\end{aligned}$$

The first three terms, $a_1L+a_2L^2+a_3L^3$, give the elastic energy of a motionless lattice while the remaining terms account for lattice vibrations. The important feature of this equation is that, even within the harmonic approximation, the prefactor of the harmonic term, (a_2+3a_3L), depends on the anharmonicity of the potential (through a_3L). In the more realistic case of three-dimensional systems, this length-dependence translates into a volume-dependence[13] of the harmonic force constants $\Phi\binom{ll'}{ij}$.

The volume dependence of the phonon frequencies induced by the volume-dependence of the force constants is traditionally modeled by the Grüneisen parameter

$$\gamma_{kj}=-\frac{\partial\ln\nu_j(k)}{\partial\ln V}$$

which is defined for each branch j and each point k in the first Brillouin zone. But since we are interested in determining the free energy of a system, it is convenient to directly parametrize the volume dependence of the free energy itself. This dependence has two sources: the change in entropy due to the change in the phonon frequencies and the elastic energy change due to the expansion of the lattice:

$$F(T,V)=E^*(V)+F_H(T,V)$$

where $E^*(V)$ is the energy of a motionless lattice constrained to remain at volume V, while $F_H(V)$ is the free energy of a harmonic system constrained to remain at volume V. The equilibrium volume $V^*(T)$ at temperature T is obtained by minimizing this quantity with respect to V. The resulting free energy at temperature T is then given by $F(T,V^*(T))$.[14]

[13] Of course, in general, it could be a general strain dependence, if the symmetry of the crystal is sufficiently low.

[14] Formally, the free energy should be determined by an sum over every possible volume: $-k_BT\ln\left(\sum_V\exp\left(-\beta F(V)\right)\right)$. However, since the volume is a macroscopic quantity, its distribution can be considered a delta function and the sum reduces to a single term: the free energy at the

Let us consider a particular case that illustrates the effect of temperature on the free energy, at the cost of a few reasonable assumptions. We assume that

- the elastic energy of the motionless lattice is quadratic in volume;
- the high temperature limit of the free energy can be used.

As shown in Appendix A.2, in this approximation, the volume expansion ΔV as a function of temperature takes on a particularly simple form:

$$\frac{\Delta V}{N} = \frac{3k_B T\overline{\gamma}}{B}$$

where $\overline{\gamma}$ is an average Grüneisen parameter:

$$\overline{\gamma} = \frac{1}{3N}\sum_{m=1}^{3N}\frac{V}{\nu_m}\frac{\partial \nu_m}{\partial V}.$$

The resulting temperature dependence of the free energy is given by

$$\frac{F(T)}{N} = \frac{F(T,V_0)}{N} - \frac{\left(3k_B T\overline{\gamma}\right)^2}{2B\left(V_0/N\right)}.$$

These expressions provide a simple way to account for thermal expansion.

Simulation

There are two main simulation-based approaches to handling anharmonicity: Monte Carlo (MC) [19] and Molecular Dynamics (MD) [5]. While both approaches are able to model anharmonicity at any level of accuracy, they suffer from two limitations. First, they are computationally demanding and therefore have, to date, been limited to simple energy models. Second, they are unable to model quantum mechanical aspects of vibrations and are therefore limited to the high temperature limit.[15]

The use of simulation techniques to determine vibrational properties bypasses the coarse-graining framework presented in section 2.2.3: Both configurational and vibrational excitations are treated on the same level. With a simple and accurate energy model, one can also calculate phase diagrams directly from MC simulations, where both atomic displacements and changes in chemical species are allowed during the simulation [116]. Obtaining phase diagrams from MD simulations would be prohibitive, because over the time scale accessible with MD, too few Ising configurations are sampled to satisfy the ergodicity requirement.

While a full determination of a phase diagram from simulations has so far not been attempted [122], both MD and MC have been used to determine differences in

volume that minimizes the free energy.

[15] Monte Carlo simulations that include quantum effects are possible for systems containing a small number of particles.

vibrational entropy between two phases. Because neither MD nor MC are able to provide free energies directly, a special integration technique has to be used. In the context of MC simulations, the procedure is called MC integration and consists in expressing a thermodynamic quantity inaccessible to MC as an integral of a quantity that *can* be obtained through MC. A simple example is the change Gibbs free energy G as function of temperature at constant pressure, which can be derived from the Gibbs-Helmholtz relation

$$F\left(T_2\right) = F\left(T_1\right) - \int_{T_1}^{T_2} \frac{E}{T^2} dT$$

where E is the internal energy.

In the context of MD simulation, free energy differences can be obtained through a process called adiabatic switching (see, for instance, [115]). This method consists in gradually changing the interatomic potentials during the course of the simulation, in order to model a change in the configuration of the alloy, without requiring atoms to jump between lattice sites. This task is achieved by defining an effective Hamiltonian

$$H_\lambda = \left(1 - \lambda\right) H^\alpha + \lambda H^\beta$$

that gradually switches from the Hamiltonian H_α associated with phase α to the Hamiltonian H_β associated with phase β as the switching parameter λ goes from 0 to 1. This convenient path of integration permits the calculation of free energy differences between phases at a reasonable computational cost, with the help of the following thermodynamic relation:

$$F^\beta = F^\alpha + \int_0^1 \left(\left\langle H_\beta \right\rangle_\lambda - \left\langle H_\alpha \right\rangle_\lambda\right) d\lambda$$

where $\left\langle H_\alpha \right\rangle_\lambda$ is the time average of the energy calculated using Hamiltonian H_α (and similarly for $\left\langle H_\beta \right\rangle_\lambda$.

There is an interesting and useful complementarity between the quasi-harmonic model and simulation techniques [46, 89]. Quantum effects typically become negligible in the temperature range where strong anharmonic effects, which cannot be modeled accurately within the quasiharmonic framework, become important.

2.4.3 Energy models

Force constants and anharmonic contributions are ultimately always derived from an energy model. In this section, we discuss various energy models, from empirical potential models to first-principles techniques, and the error or bias they may introduce in the vibrational properties.

Simple pairwise potentials of functionals (such as the Embedded Atom Method) are computationally efficient so that all vibrational properties can often be determined without any approximations beyond the ones associated with the specific energy model. For this reason, the use of simple energy models has proven to be an

T (K)	$S_{vib}^{o\rightarrow d}$	$\Delta V^{o\rightarrow d}/V^{o}$ (%)	Reference
high	0.29	3%	[2]
1200	0.22	2%	[115]
1000	0.15	2%	[8][9]
1000	0.11	1.6%	[89], Foiles-Daw EAM potentials
high	0.00	0.5%	[136]

Table 2.4: Relation Between the Vibrational Entropy Change upon Disordering and the Volume Change upon Disordering in Various Theoretical Investigations of the Ni_3Al Compounds

invaluable tool to understand trends in vibrational entropies and to test a number of approximations [54, 46, 90, 55, 89].

Several potential sources of error can arise when using pair potentials or pair functionals. The first one is that vibrational entropy is extremely sensitive to the precise nature of the relaxations that take place in an alloy and a simple energy model may not be able to accurately predict these relaxations. This problem is particularly apparent when considering the wide range of values found in the different calculations of the vibrational entropy change upon disordering of the Ni_3Al compound [2, 8, 115, 136]. But, as shown in Table 2.4, most of the discrepancies can be explained from differences in the predicted volume change upon disordering.

This is often aggravated by the fact that simple energy models are often not fitted to phonon properties. The problem was noted in [8] where the embedded atom potentials used were fitted to various structural energies and elastic constants [140]. The acoustic modes were accurately extrapolated from the fit to the elastic constants, but the phonon frequencies associated with the optical modes were overestimated by about 10%.[16] The question of the accuracy of simple energy models clearly merits further attention. In this respect, the fit of simple energy models to the results of *ab-initio* calculations [122, 121] offers a promising way to include vibrational effects.

In oxides, electronic polarization has to be included in order to correctly model both the low frequency acoustic modes and the high frequency optical modes. Electronic polarization in oxides can be included with the so-called core and shell model.

While quantum mechanical methods are computationally more intensive, they generally provide more accurate force constants. The most obvious error introduced by the common Local Density Approximation (LDA) is its systematic underprediction of lattice constants which leads to an overestimation of elastic constants and phonon frequencies. This systematic error makes it difficult to compare the absolute values of calculated vibrational properties with experimental measurements. However, for the purpose of calculating phase diagrams, this bias may be less of a concern, because

[16]Most of the bias in the vibrational entropy introduced by this problem should however cancel out when taking the difference in vibrational entropy between two phases where the same problem is present.

phase stability is determined by differences in free energies, and one would expect a large part of this systematic error to cancel out.

A practical way to alleviate the LDA bias is to perform calculations at a negative pressure such that the calculated equilibrium volume agrees with the experimentally observed volume. As shown in [135], a very good estimate of the required negative pressure can be obtained by a concentration-weighted average of the pressure associated with the elemental solids. For the purpose of calculating elastic properties, this approach appears to outperform the most popular alternative to LDA, the Generalized Gradient Approximation (GGA).[17]

2.5 Experimental techniques

The experimental literature on the thermodynamics of lattice vibrations in alloys relies on mainly three techniques.

In *differential calorimetry* measurements, the heat capacity of two samples in a different state of order is compared over a range of temperatures. If the upper limit of the range of temperatures is chosen to be sufficiently low, substitutional exchanges will not occur and the difference in heat capacity can be assumed to arise solely from vibrational effects. Integration of the difference in heat capacity (divided by temperature) then yields a direct measure of the vibrational entropy differences between the two samples of the range of temperature considered. This, of course, assumes that the lower temperature bound is sufficiently low, so that the vibrational entropy of both samples can be assumed to be zero at that temperature. It also assumes that the electronic contribution to the heat capacity is negligible. In practice, both assumptions are typically satisfied. The main problem with this method is that one is usually interested in vibrational entropy differences at the transition temperature of the alloy, which is usually above the upper limit of the temperature range used in the heat capacity measurements. The heat capacity therefore needs to be extrapolated to high temperature. This constitutes the main source of inaccuracies in this method. Examples of the use of this method can be found in [11, 10, 96, 95, 98, 97].

A second method is the measurement of phonon dispersion curve through *inelastic neutron scattering*. For ordered alloys that can be produced in large single crystals, this method is very powerful. Once the dispersion curves along special directions in reciprocal space are measured, they can be used to fit Born-von Kármán spring models which, in turn, yield the normal frequencies for any point in the Brillouin zone. With the help of the standard statistical mechanics techniques described in Section 2.4.1, this information is sufficient to determine the vibrational entropy. Examples of applications of this method can be found in [10, 96, 50, 49, 95]. The applicability of this method is unfortunately limited by the availability of large single crystals. The case of disordered alloys presents an even more fundamental problem: Disordered

[17]Part of the success of the "negative pressure" LDA is due to the fact that it uses information regarding the true experimental volume instead of being fully ab-initio. But the knowledge of one pressure per element is a relatively small amount of information.

alloys do not have well defined dispersion curves and there is no straightforward way to fit the spring constants of a spring model from the experimental data. This problem is usually addressed by using the virtual crystal approximation, in which different constituent atoms are replaced by one "average" type of atom (see Appendix A.3). Unfortunately, this approximation has repeatedly been shown to have a very limited accuracy for the purpose of measuring vibrational entropy differences [8, 115, 121, 96]. Nevertheless, single crystal phonon dispersion curve measurements for ordered alloys present a unique opportunity to perform a stringent test of the accuracy of theoretical models.

A third method is the determination of the phonon density of states from *incoherent neutron scattering* measurements. In contrast to the preceding approach, this method can readily be applied to disordered systems and to compounds for which single crystals are not available [49, 98, 97, 21, 117]. The main limitation of this approach is that different atomic species have different neutron scattering cross-sections. The scattered intensity at each frequency measures a "density of states", where each mode is weighted by the scattering intensity of the atoms participating in the mode in question. Thus, one needs some prior information about the vibrational modes in order to reconstruct the true phonon DOS from the experimental data. In the case of alloys, there is not a one-to-one correspondence between the measured data and the vibrational entropy. This problem can be alleviated by choosing alloy systems where the scattering intensity of each species is similar [97, 117].

Other techniques have been used to measure vibrational entropy differences. Some researchers have used the fact that vibrational entropy and thermal expansion are directly related to estimate vibrational entropy differences from accurate thermal expansion measurements [93, 94]. The measurement of inelastic nuclear resonant scattering spectrum has also been used to relate changes in the phonon DOS to changes in the short-range order of a disordered alloy [52]. Finally, relatively noisy estimates of vibrational entropy differences can be obtained from X-ray Debye-Waller factors or from the measurement of mean square relative displacement (MSRD) of the atoms relative to their neighbors through extended electron energy-loss fine structure (EXELFS) [11].

2.6 The origin of vibrational entropy differences between phases

We have presented the framework that allows for the inclusion of vibrational effects in phase diagram calculations. However, the formalism presented so far does not directly provide any intuition regarding the origin of vibrational entropy differences. This intuition is important to be able the predict when vibrational effects should be important and, when they are, which approximation should be used to calculate them.

Three mechanisms have been suggested to explain the origin of vibrational entropy differences in alloys. We will discuss them in turn.

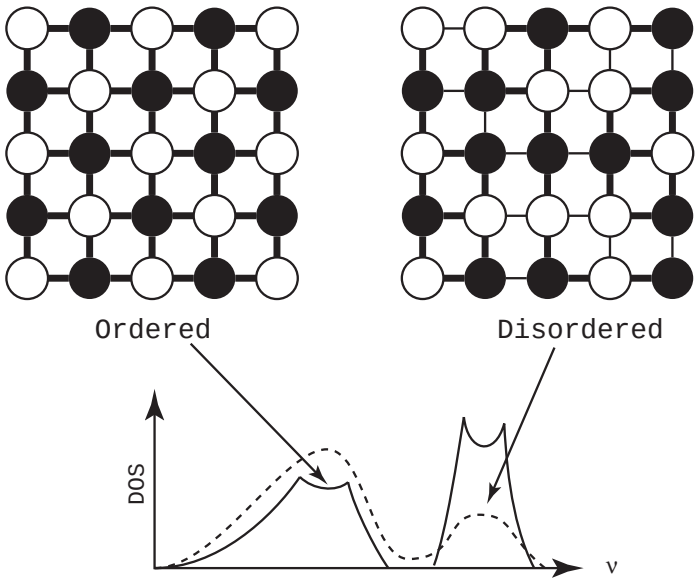


Figure 2-2: The "Bond Proportion" Mechanism.

2.6.1 The "bond proportion" effect

In most theoretical studies based on simple models systems [82, 14, 16, 133, 53], the effect of the state of order of an alloy on its vibrational entropy has been attributed to the fact that bonds between different chemical species have a different stiffness than the bonds between identical species. When the proportion of each type of bond in the alloy changes during, for instance, an order-disorder transition, the average stiffness of the alloy changes as well, resulting in a change of its vibrational entropy. This so-called "bond-proportion" mechanism is illustrated in Figure 2-2 in the case of an order-disorder transition. In a system with ordering tendencies, the bond between unlike atoms are associated with an increased stability and are thus expected to be stiffer than bonds between alike atoms. Since disordering reduces the number of bonds between unlike atoms in favor of bonds between similar atoms, the disordered state is expected to be softer, and thus, have a large vibrational entropy. Vibrations would then tend to stabilize the disordered state relative to the ordered state, reducing the transition temperature. A similar reasoning in the case of a phase separating system shows that the disordered state should be softer than a phase separated mixture, indicating that the miscibility gap should be lowered as a result of vibrational effects.

The presence of a "bond proportion" effect can be readily identified from the nature of the changes taking place in the phonon densities of states during an order-disorder transition. In the ordered alloy, the very stiff nearest-neighbor bonds should be associated with high frequency optical modes peaks. As the alloy disorders, the height of these peaks should decrease since the number of stiff bonds decreases, This characteristic signature of the "bond proportion" mechanism in the phonon DOS has been repeatedly observed in experiments [49, 10].

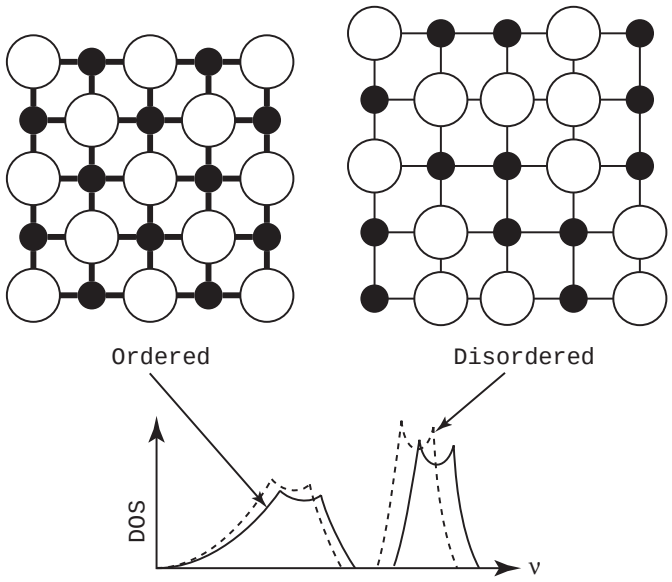


Figure 2-3: The Volume Mechanism.

2.6.2 The volume effect

It is well known that vibrational entropy of a given compound varies with volume: this dependence is responsible for thermal expansion. It is thus expected that the volume change that typically occurs during solid-state transitions should also result in a change in vibrational entropy. As the alloy expands (or contracts), as a result of a change in its state of order, the stiffness of all chemical bonds decreases (or increases). The resulting change in vibrational entropy is entirely due to anharmonicity, in contrast to the "bond proportion" effect. In the limit of a "pure" volume effect, the spatial distribution of the stiffness of each bond is inconsequential: The state of order of the alloy influences its vibrational entropy only through a change in overall volume. When the volume mechanism operates alone, the phonon DOS should exhibit an overall shift when the volume changes (See Figure 2-3).

This shift is usually accompanied by a change in the shape of the phonon DOS, so that visual inspection of the phonon DOS is usually not sufficient to identify the effect of volume. Theoretical investigations of this effect thus rely on a simple thought experiment consisting in separating the vibrational entropy change upon disordering at constant volume from the vibrational entropy change resulting solely from the volume expansion of the disordered state. In this fashion, the importance of volume changes was first observed in embedded atom method (EAM) calculations of disordering reaction of the Ni_3Al and Cu_3Au compounds [2] and later corroborated by subsequent EAM calculations on the Ni_3Al compound [8, 115]. These more recent calculations also found that the volume effect is magnified by the fact that the linear thermal expansion coefficient of different phases can differ substantially (by about $5 \times 10^{-6}\ K^{-1}$ relative to an absolute value of about $15 \times 10^{-6}\ K^{-1}$). First-principles calculations on the Cu-Au system [106] revealed a similar finding. In contrast, first-principles calculations on the Ni_3Al [136] and Pd_3V [137] compounds found only a

small difference between the thermal expansion coefficients of the ordered and the disordered phases (less than $1 \times 10^{-6}\ K^{-1}$).

Even though the temperature-dependence of vibrational entropy can be large, it is important to keep in mind that a given vibrational entropy difference arising from anharmonic contributions has a smaller impact on the vibrational free energy than a harmonic contribution of the same magnitude. The reason is that anharmonic contributions to the vibrational entropy are always partly cancelled by anharmonic contributions to the vibrational *enthalpy*. In contrast, such cancellation does not occur for harmonic contributions.[18] In its simplest form, the quasiharmonic approximation predicts that anharmonic contribution have exactly half the effect of harmonic contributions (see [106] or Appendix A.2).

Experimental measurements have not identified the volume mechanism as a major source of entropy differences since the simple thought experiment that allows its identification in calculations cannot be performed experimentally. The effect of thermal expansion on the phonon DOS, however, is clearly seen experimentally [98, 97], due to the fact that thermal expansion causes shifts in the phonon DOS that are not accompanied by substantial changes in its shape. Measurements on Ni_3V [98], Co_3V [97] and Cu_3Au [21] all show that anharmonic contributions are not negligible.

2.6.3 The size mismatch effect

The third advocated source of vibrational entropy changes is the effect of atomic size mismatch. When atoms of different sizes are constrained to coexist on a lattice, the atoms can experience compressive (or tensile) stress that results in locally stiffer (or softer) regions. When large atoms sit on neighboring lattice sites, the amplitude of their vibrations is reduced, *i.e.*, the alloy tends to be locally stiffer. Conversely, when small atoms sit on neighboring lattice sites, the extra room available results in a locally softer region.

The phenomenon was first noted in EAM calculations on the Cu_3Au compound [2], where, in the disordered state, the presence of highly compressed pairs of Au atoms lowers the vibrational entropy of the disordered state. A similar effect was found in first-principles calculations on Ni_3Al [136], where very compressed pairs of Al atoms where found in the disordered state. In first-principles calculations on Pd_3V, an even more intriguing size-related effect was observed: all three types of chemical bonds have incompatible equilibrium lengths and the vibrational entropy changes can be explained solely by the large relaxations of the atoms away from their ideal lattice sites in the disordered state.

[18] A temperature-dependence of vibrational entropy necessarily introduces a temperature-dependence of the vibrational enthalpy, as a consequence of the following thermodynamic relation: $\frac{\partial H_{vib}}{\partial T} = T\frac{\partial S_{vib}}{\partial T}$.The anharmonic vibrational free energy is then a sum of two competing contributions: $F_{vib} = H_{vib} - TS_{vib}$. In contrast, harmonic contributions to the vibrational enthalpy are configuration-independent in the high-temperature limit, by the equipartition theorem, and give no net contribution to vibrational free energy differences. Vibrational entropy differences originating from harmonic contributions thus enter the expression for vibrational free energy differences directly, without any partial cancellation from the enthalpic term.

In the experimental literature, the size mismatch effect has been described with the help of a stiff sphere picture first introduced in an investigation of the Cu_3Au compound [95]. The fundamental intuition behind this picture is that a chemical bond becomes stiffer when the two bonded atoms have touching atomic "spheres". Further evidence for this stiff sphere picture was provided by a systematic analysis of the vibrational entropy of formation of various $L1_2$ compounds [20], which shows a correlation with the difference in radii between the two alloyed species.

2.7 Controlled approximations

The basic formalisms presented in Section 2.2.1 and 2.4.1 provide two natural ways to control the trade-off between accuracy and computational requirements. In the context of alloy theory (Section 2.2.1), the range of the effective clusters interactions included in the cluster expansion controls how accurately the configurational dependence of vibrational properties is modeled. In the context of the harmonic (or quasiharmonic) treatment of lattice vibrations (Section 2.4.1), the range of the force constants included in the Born-von Kármán model controls the accuracy of the calculated vibrational properties for a given configuration. In principle, any desired accuracy can be reached, given sufficient computing power, by increasing the range of the interactions in both the cluster expansion and the Born-von Kármán models. This section seeks to answer the important question of how far these two ranges of interactions need to be pushed in order to reach the accuracy required in a typical phase diagram calculation.

2.7.1 Short-range force constant

The evidence that spring models including only short-range force constants are able to correctly model vibrational quantities comes from various sources.

First and second nearest neighbor spring model are routinely used to fit data obtained from neutron scattering measurement of phonon dispersion curves [49, 95]. In the theoretical literature, there have been direct studies of the convergence as a function of the range of interaction considered. All *ab-initio* studies find that short-range force constants (first or second nearest neighbor) permit an accurate determination of thermodynamical quantities in metals [136, 137] and group IV semiconductors [55]. It is important to note that this rapid convergence of most thermodynamic quantities occurs even when the pointwise convergence rate of the phonon DOS is slow. As noted before, this property arises from the fact that thermodynamic quantities are averages taken over all phonon modes and errors tend to average out.

In ionic systems, the presence of long-range electrostatic interactions may require long-range force constants. However, this electrostatic effect can easily be modeled using pair potentials at a moderate computational cost. Once the forces predicted from a simple electrostatic model have been subtracted, the residual forces should be parameterizable with a short-range spring model.

Some of the ab-initio studies of convergence have suggested additional simplifications to force constant tensors [55, 137]: instead of attempting to compute all force constants in each tensor, is it possible to obtain reliable results by keeping only the largest terms. We now present a hierarchy of approximations that is a formalization of these findings.

To obtain a more intuitive representation of a given force constant tensor $\Phi_{\alpha\beta}(i,j)$, we express it in a basis such that the first cartesian axis is aligned along the line joining atom i and j. The second axis is then taken along the highest symmetry direction orthogonal to the first axis while the third axis is chosen so obtain a right handed orthogonal coordinate system.

In the absence of symmetry, the most general force constant tensor has 9 independent elements. The first simplification, is to neglect the three body terms in the harmonic model of the energy (*e.g.* $(x_\alpha(i) - x_\alpha(j))(x_\beta(i) - x_\beta(k))$ with $\alpha \neq \beta$). Physically, such terms arise from the deformation of the electronic cloud surrounding atom i that is caused by moving atom j and that affect the force acting on atom k. Clearly, for any force constant other than the nearest neighbor, this effect is negligibly small. Even for nearest neighbor tensors, it is the most natural contribution to neglect first. In can be readily shown that a solid consisting only of pairwise harmonic interaction, the tensor associated with a pair of atoms is symmetric:

$$\Phi_{\alpha\beta}(i,j) = \Phi_{\beta\alpha}(i,j).$$

(This constraint is distinct from the conventional constraint: $\Phi_{\alpha\beta}(i,j) = \Phi_{\beta\alpha}(j,i)$.)

The elements of the force constant tensor can be ranked in decreasing order of expected magnitude based on three simple assumptions:

1. Force constants associated with stretching a bond are larger than the ones associated with bending it.

2. Terms relating orthogonal forces and displacements are smaller than those relating parallel forces and displacements.

3. In the plane perpendicular to the bond, the anisotropy in the force constants is smaller than the magnitude of the force constants themselves.

We then obtain

$$\Phi(i,j) = \begin{pmatrix} a & d+e & d-e \\ d+e & b+c & f \\ d-e & f & b-c \end{pmatrix}$$

with

$$a > b > c > d > e > f.$$

This hierarchy of force constants is important to keep in mind, given that the off-diagonal elements of the spring tensors are the most difficult to obtain from supercell

calculations, requiring much bigger supercells than diagonal elements. There is evidence [137] that even keeping only the stretching (a) and isotropic bending (b) terms of the nearest neighbor spring tensor can provide an accuracy of about 0.03 k_B. If this observation turns out to be generally applicable, this offers a simple way to account for vibrational effects in phase diagram calculations.

2.7.2 Short-range effective cluster interactions

If a cluster expansion of the vibrational free energy only requires a small number of ECI to accurately model the configurational-dependence of the vibrational free energy, it then becomes practical to determine the values of these ECI from a small number of very accurate calculations of the vibrational free energy of a few structures.

The issue of the speed of convergence of the cluster expansion is also related to the task of devising efficient ways to compute vibrational properties of disordered alloys: The faster the cluster expansion converges, the easier it is to model a disordered phase (see Appendix A.3). The calculations of the vibrational entropy change upon disordering has proven to be a very effective way to assess the importance of lattice vibrations [8, 115, 136, 137], since this quantity can be straightforwardly used to estimate the effect of lattice vibrations on transition temperatures with the help of Equation (2.4).

The central question is thus whether the cluster expansion of the vibrational free energy converges quickly with respect to the number of ECI. This is a question distinct from the range of force constants needed to obtain accurate vibrational properties. The range of ECI needed to represent the configurational dependence of vibrational free energy may very well exceed the range of the force constants. Even in simple Born von Kármán model systems, there is no direct correspondence between ECI and force constants, except in special cases (see Section 2.8.1). Once relaxations are introduced in the model, then all hope of a simple correspondence is lost [90].

In this context, the question of the existence of a rapidly converging cluster expansion of vibrational properties has to be answered through numerical experiments. Simple energy models offer the possibility to test, at a reasonable computational cost, the speed of convergence of a cluster expansion. Explicit calculations of a well converged cluster expansion of vibrational entropy in a Lennard-Jones solid [54] have indicated that a small number of ECI (9) can provide a good accuracy ($\pm 0.03 k_B$). Other benchmarks of the speed of convergence, based on studies of disordered alloys [46, 90, 89, 91], also indicate that concise and accurate cluster expansions are possible. Experiments that seek to link features of projected phonon DOS to the local chemical environment of the atoms [52] suggest that short-range ECI should be able to successfully model vibrational entropy differences. One potential source of concern is the difficulty associated with accounting for the size mismatch effect using a short-range ECI [90, 137]. In the context of cluster expansions of the energy, relaxations of the atoms away from their ideal lattice site as a result of size mismatch are known to introduce both non negligible long range pair ECI and numerous multiplet ECI. A cluster expansion of the vibrational free energy is expected to exhibit the same problems.

All the full phonon DOS ab-initio calculations of vibrational entropies in alloy systems performed so far have relied on the rapid convergence of the cluster expansion [55, 129, 136, 106, 137]. While efforts to quantify the error introduced by truncating the cluster expansion in ab-initio calculations have been made [55, 136, 137], the issue of the speed of convergence of the cluster expansion in the context of vibrational properties clearly merits further study, especially in light of the importance of the size mismatch effect [137].

2.8 Models of lattice vibrations

While the ability to control the level of approximation discussed in the previous section is extremely useful, there remains the problem that, very often, only considering the first few levels in this hierarchy of approximations already involves substantial computational requirements. For this reason, models of lattice vibrations that involve fewer parameters but more physical intuition may provide a practical mean of including vibrational effects in phase diagram calculations. In this section, we will present the advantages and weaknesses of each method, in light of the three fundamental mechanisms described in the Section 2.6.

2.8.1 The "bond proportion" model

There have been many attempts (see, for instance, [42, 144, 14, 53]) to find ways to express the relationship between the vibrational free energy and the dynamical matrix in a form that illustrates the intuition behind the "bond proportion" mechanism. For simple nearest-neighbor spring models with central forces, in linear chains [14, 82, 53], square [14] or simple cubic lattices [141], a convenient exact expression can be derived for the nearest neighbor ECI in the expansion of the vibrational free energy in the high temperature limit:

$$V_{1nn} = \frac{d}{8} k_B T \ln \left(\frac{k_{AA} k_{BB}}{k_{AB}^2} \right). \tag{2.14}$$

where k_{AA}, k_{BB} and k_{AB} are, respectively, the spring constants associated with $A-A$, $B-B$ and $A-B$ bonds and d is the dimensionality of the system. It has been noted, on the basis of numerical experiments, that the same expression performs well for other lattices [53]. This success arises from the fact that, as shown in Appendix A.5, Equation (2.14) is the first order approximation to the true vibrational entropy change in a large class of systems which satisfies the following assumptions:

- the high temperature limit of the vibrational entropy is appropriate;
- the nearest-neighbor force constants can be written as $\Phi(i,j) = k_{\sigma_i \sigma_j} \phi(i,j)$ where $k_{\sigma_i \sigma_j}$ denotes the (scalar) stiffness of the spring connecting sites i and j with occupations σ_i and σ_j while the $\phi(i,j)$ are dimensionless spring constant tensors. The $\phi(i,j)$ are assumed equivalent under a symmetry operation of the space group of the parent lattice;

- all force constants $k_{\sigma_i\sigma_j}$ are such that

$$\left| \frac{k_{\sigma_i\sigma_j}}{\sqrt{k_{\sigma_i\sigma_i}k_{\sigma_j\sigma_j}}} - 1 \right| \ll 1. \tag{2.15}$$

Equation (2.14) applies to simple harmonic models with nearest neighbor springs on the fcc, bcc or sc primitive lattices (and, approximately, on the hcp lattice), as long as the above assumptions are satisfied. Both stretching and bending terms are allowed in the spring tensors, as long as their relative magnitude is independent of $k_{\sigma_i\sigma_j}$ (*e.g.* when the bending terms are always, say, 10% of the corresponding stretching term, regardless of the magnitude of the stretching term). Why focus only on stretching and bending terms, even though the spring tensor associated with a pair of atoms can contain up to 9 independent terms? The "bond proportion" picture requires every bond of a certain type (for instance, $A-A$ bonds) to be have an identical spring tensor. However, the point symmetry of each bond can be different and similar chemical bonds in different environment face different symmetry-induced constraints on their spring tensors [123]. The only way to reconcile these observations is use a spring tensor that is compatible with the highest possible symmetry, ensuring that it is also compatible with any other environment with a lower symmetry With the highest possible symmetry, only two independent terms remain in the spring tensor: the stretching and bending terms.

Equation (2.14) embodies the essential intuition behind the effect of the alloy's state of order on its vibrational free energy in the bond proportion approximation. When one replaces a $A-A$ bond and a $B-B$ bond by two $A-B$ bonds, the vibrational free energy will decrease only if the stiffness of $A-B$ bonds, k_{AB}, exceeds the geometrical average stiffness of the bonds between identical species $\sqrt{k_{AA}k_{BB}}$. This observation allows the determination of the expected effect of vibrations on the shape of the phase diagram by simple arguments. The link between the nearest neighbor ECI of the expansion of the vibrational entropy can be summarized by the expression [54]:

$$\frac{T_c^{\text{config+vib}}}{T_c^{\text{config}}} = \frac{1}{1 \mp \alpha V_{1nn}/k_BT}$$

where the "$-$" and "$+$" correspond to ordering and segregating systems, respectively, and where α is a dimensionless parameter that only depends on the lattice type and the ordering tendency of the system (for instance, for fcc, $\alpha = 1.7$ in ordering systems and $\alpha = 9.8$ in segregating systems, while for bcc, $\alpha = 6.5$ in both cases).

It is very simple to include vibrational effects in phase diagram calculations using the "bond proportion" model. All that is needed is an estimate of the stiffness of $A-A$, $B-B$ and $A-B$ bonds, which could come, for instance, from supercell calculations of the nearest neighbor force constants in a few simple structures or from the bulk moduli of the pure elements and one ordered compound. The nearest neighbor ECI then obtained can be simply added to the cluster expansion of the

energy.

While Equation (2.14) is useful to estimate the importance of the "bond proportion" mechanism in a given system, one can avoid some of the approximations involved in deriving Equation (2.14) at the expense of only a modest amount of additional computing power. One can find the exact phonon DOS of the nearest neighbor Born-von Kármán model for a variety of configurations of the alloy, which allows a more accurate cluster expansion of the vibrational energy to be derived. In this fashion, the condition specified in Equation (2.15) is no longer needed and the vibrational entropy can be calculated at any temperature.

It is important to keep in mind that two important assumptions are made when invoking the "bond proportion" mechanism. First, vibrational entropies are solely determined by the nearest neighbor force constants. There is theoretical evidence that nearest neighbor spring models can predict vibrational entropy differences with an accuracy of about $0.02k_B$ in metallic [136, 137] and semiconductor [55] systems. Given that configurational entropy differences are typically of the order of $0.5k_B$, this precision should be sufficient for practical phase diagram calculations.

The second assumption is that each type of chemical bond is assumed to have an intrinsic stiffness that is independent of its environment. First-principles calculations on the Li-Al [123] and on the Pd-V system [137] unfortunately indicate that the stiffness of a chemical bond does change substantially as a function of its environment. This problem is serious, as it considerably limits the applicability of the "bond proportion" model. These changes of the intrinsic stiffness of the bonds as a function of their environment are precisely the focus of the two other suggested sources of vibrational entropy changes. In summary, while the "bond proportion" model gives an elegant description of one of the mechanisms suggested to be at the origin of vibrational entropy differences, it completely ignores the two other mechanisms, namely, the volume and size mismatch effects.

2.8.2 The Debye model

Perhaps the most widespread approximation to the phonon DOS $g(\nu)$ is the Debye model [58], where the phonon problem is solved in the acoustic limit. In this case, the phonon DOS is approximated by:

$$g(\nu) = \begin{cases} \frac{9\nu^2}{\nu_D^3} \text{ if } \nu \leq \nu_D \\ 0 \text{ if } \nu > \nu_D \end{cases}$$

where $\nu_D = \frac{k_B \Theta_D}{h}$ and Θ_D is the Debye temperature, given by:

$$\Theta_D = \frac{h}{k_B} \left(\frac{3N}{4\pi V} \right)^{1/3} C_D$$

where C_D is the Debye sound velocity, defined by

$$\frac{3}{C_D^3} = \left\langle \sum_{\lambda=1}^{3} \frac{1}{C_\lambda^3} \right\rangle$$

where the right-hand side is the directional average of a function of the three sound velocities C_λ [58]. The free energy of a Debye solid is given by:

$$\begin{aligned} \frac{F}{N} &= \frac{E^*}{N} + \frac{9}{8} k_B \Theta_D - k_B T \left(D\left(\frac{\Theta_D}{T}\right) - 3\ln\left(1 - \exp\left(-\frac{\Theta_D}{T}\right)\right)\right) \\ &\approx \frac{E^*}{N} + 3k_B T \ln\left(\frac{k\Theta_D}{h}\right) \text{ in the high temperature limit.} \end{aligned}$$

where the Debye function $D(u)$ is given by

$$D(u) = 3u^3 \int_0^u \frac{x^4 e^x}{(e^x - 1)^2} dx$$

Since the Debye sound velocity C_D is a complicated function of all elastic constants of the material, an approximation to the Debye temperature that only involves the bulk modulus proves extremely useful. Such an approximation was derived by Moruzzi, Janak and Schwarz (MJS) [92] for cubic materials[19]:

$$\Theta_D = 0.617 \left(\frac{3}{4\pi}\right)^{1/3} \frac{h}{k_B} \left(\frac{\Omega^{1/3} B}{\overline{M}}\right)^{1/2}$$

where Ω is the average atomic volume B is the bulk modulus and $\overline{M}$ is the concentration weighted arithmetic mean of the atomic masses. As noted in [53], in the high temperature limit, the MJS model does not exhibit the property that the masses have no effect on the vibrational free energy of formation, although using a geometric average of the masses [54] fixes this problem.

The quasiharmonic approximation can be used, within Debye theory, to account for mild anharmonicity. In the so-called Debye-Grüneisen approximation, the volume-dependence of the phonon DOS is modeled by a single Grüneisen parameter and the effect of volume can be summarized by simply making the Debye temperature volume-dependent:

$$\Theta_D = \Theta_{D,0} \left(\frac{V_0}{V}\right)^{\gamma}.$$

where $\Theta_{D,0}$ is the Debye temperature at volume V_0 and γ is the grüneisen parameter.

Despite of its inaccurate description of the true phonon DOS at high frequencies, the Debye and Debye-Grüneisen models are quite successful at modeling the changes

[19] although it has been used for materials with a lower symmetry [13, 120]

in vibrational properties of a given compound as a function of temperature. For instance, the thermal properties of pure metals [92] calculated in MJS approximation are surprisingly accurate. The reason for this success is that the distinctive trends of most thermodynamic quantities occur at low temperature, where the low frequency phonon modes that are correctly described by the Debye model have a dominant effect. In the high temperature regime, thermodynamic quantities are determined by the classical equipartition theorem, and any harmonic model gives the correct behavior.

Debye-like models are expected to perform well in systems where the differences in vibrational free energy between compounds can be explained by uniform shifts of the phonon DOS, such as when the volume effect operates alone. Such a behavior has been observed in Embedded Atom calculations on the Ni–Al system [2, 8, 115] but in no other systems so far. The MJS approximation has been used to include vibrational effects in phase diagram calculations and has resulted in an improved agreement with experimental results [13, 120, 34].

However, as shown in [54], the Debye approximation and its successors can have significant shortcomings when used to calculate phase diagrams. A significant part of the vibrational free energy differences between different compounds arises from changes in the high frequency portion of the phonon DOS, which Debye-like models describe incorrectly. In some cases, the MJS approximation can even lead to an incorrect prediction of the sign of the vibrational entropy difference [137, 106].

In summary, the Debye model and its derivatives capture the essential physics behind only one of the advocated mechanisms responsible for the configurational dependence of vibrational free energy: the volume effect. Approximations based on the Debye model, however, fail to account for the possibility that the state of order also has a direct impact on the shape of the phonon DOS (as predicted, for instance, by the "bond proportion" model), especially in the high frequency portion of the phonon DOS.

2.8.3 The Einstein model

At the other extreme of the spectrum is the Einstein approximation, which focuses on the high frequency region of the phonon DOS and therefore has the potential of better representing the effect of the state of order on vibrational entropy [53].

The Einstein model can be combined with a Debye model to better fit experimental calorimetry data [10] or thermal expansion data [93]. The Einstein model, also called the local harmonic model [78, 128], has proven especially useful to calculate vibrational entropies associated with defects [146, 99, 128].

In quantitative theoretical calculations, this model may not have many applications, because determining unknown parameters of an Einstein model from first-principles is as difficult as obtaining force constants, which would allow for a more precise description of the DOS to be used instead.

The Einstein model is nevertheless extremely useful for conceptual purposes, as we will now illustrate. As shown in Appendix A.4, the vibrational free energy of a system is bounded by above and by below by the free energy of two Einstein-like

systems:

$$\frac{k_B T}{2} \ln\left(\frac{h}{k_B T}\prod_i M_i \left(\left(\Phi^{-1}\right)_{ii}\right)^{-1}\right) \leq F_{\text{vib}} \leq \frac{k_B T}{2} \ln\left(\frac{h}{k_B T}\prod_i M_i \Phi_{ii}\right)$$

While the upper bound is obtained from the usual Einstein model, where surrounding atoms do not relax, the lower bound is obtained when the surrounding atoms are allowed to relax freely. Another way to interpret these bound is that at one extreme, each atom sees the others as having an infinite mass, while at the other extreme, each atom sees the other atoms as being massless. This result supports the view that vibrational free energy can be meaningfully considered as a measure of the average stiffness of each atom's local environment.

A more rigorous way of defining the contribution of an atom to the total vibrational free energy is the use of the projected DOS (also called partial DOS). This approach does not in any way simplify the calculation of vibrational properties, because the full phonon DOS is needed as an input, but it is a useful way to interpret the experimentally measured or calculated phonon DOS. To obtain the contribution of atom i to the DOS, the idea is to weight each normal mode by the magnitude of the vibration of atom i:

$$g_i(\nu) = \frac{1}{N}\sum_j \left|e_j(i)\right|^2 \delta(\nu - \nu_i)$$

where e_j is the eigenvector (normalized to unit length) associated with the mode of frequency ν_j. Since extensive thermodynamic properties are linear in the DOS, atom-specific local thermodynamic properties can be readily defined from the projected DOS. Note that, by construction, all the projected DOS sum up to the true phonon DOS and thus, all the local extensive thermodynamic quantities sum up to the corresponding total quantity.

2.8.4 The "stiffness vs. length" approach

In *ab-initio* calculations, most of the computational burden comes from the calculation of the force constant tensors. It would thus be extremely helpful if the force constants determined in one structure could be used to predict force constants in another structure. From the failures of the "bond proportion" model, however, we know that forces constants obtained from one structure are not directly transferable to another structure [123, 137]

Nevertheless, a simple modification of the transferable force constant approach yields substantial improvements in precision. First-principles calculation the Pd-V [137] system revealed that most of the variation in the stiffness of a given chemical bond across different structures can be explained by changes in bond length.[20] This

[20]The results obtained in the Li-Al system [123] also suggest than length is a good predictor of stiffness, although this matter was not further investigated.

Compound (Structure)	"Stiffness vs. Length" Model	1nn Spring Model
Pd_3V $L1_2$	-4.39	-4.39
Pd_3V $D0_{22}$	-4.42	-4.47
Pd_3V SQS-8	-4.56	-4.54
Ni_3Al $L1_2$	-5.57	-5.55
Ni_3Al SQS-8	-5.54	-5.57

Table 2.5: Comparison Between Vibrational Entropies Obtained from the "Stiffness vs. Length" Model and from a First Nearest Neighbor Spring Model.

suggests that the transferable quantity to consider is a "bond stiffness" vs. "bond length" relationship. As a first approximation, a linear relationship can be used

$$\begin{aligned} a(l) &= a_0 + a_1 (l - l_0) \\ b(l) &= b_0 + b_1 (l - l_0) \end{aligned}$$

where a and b denote the stretching and isotropic bending terms, respectively and where a_0 and b_0 describe the stiffness of the bond at its ideal length l_0 while a_1 and b_1 are analogous to bond-specific Grüneisen parameters.

This approximation was shown to be successful in the Pd-V system [137]. Figure 3-2 illustrates the ability of this simple model to predict bond stiffness in different structures. A similar analysis performed with the data on the Ni-Al system from Reference [136] is shown in Figure 3-10. Table 2.5 compares the predictions obtained from the "stiffness vs. length" model with more accurate calculations.

There are numerous advantages to this approach. From conceptual point of view, this model presents a concise way to summarize all three mechanism suggested to be the source of vibrational entropy differences. The "bond proportion" mechanism is the particular case obtained when little changes in bond length occur. The volume effect results from expanding all bonds by the same factor. The size mismatch effect (or the "stiff sphere" picture) is also modeled since the local change in stiffness resulting from locally compressed or expanded regions are explicitly taken into account. A straightforward way to represent the source of vibrational changes is to overlap the bond stiffness vs. bond length relationship and the changes in average length and stiffness in different states of order, as shown in Figure 3-4.

A second advantage of this method is computational efficiency. The unknown parameters of the model can be determined by a small number of supercell or linear response calculations. After that, the knowledge of the relaxed geometry of a structure is sufficient to determine the stiffness of all chemical bonds. Finding the vibrational entropy of the structure then just reduces to a computationally inexpensive Born-von Kármán phonon problem with nearest-neighbor force constants. It is important to note that the knowledge of the relaxed geometries of a set of structures is a natural by-product of first-principles calculations of structural energies, which are needed to construct the cluster expansion of the energy in any phase diagram calculations,

whether vibrational effects are included or not. Since computational requirements do not grow rapidly with the number of structures considered, this opens the way for a much more accurate representation of the configurational-dependence of the vibrational free energy.

An third advantage of transferable stiffness vs. length relationships is that they contain all the information needed to account for thermal expansion as well, within the quasi-harmonic approximation. The slopes of the stiffness vs. length relationships for each chemical bonds explicitly defines the changes in phonon frequencies as volume changes. Since the bulk modulus of each structure is also a by-product of structural energy calculations, all the ingredients needed for a quasi-harmonic treatment of thermal expansion are available.

2.9 Conclusion

The standard framework of alloy theory can be straightforwardly extended to account for lattice vibrations using the concept of coarse-graining of the partition function. Once the degrees of freedom associated with lattice vibrations are integrated out, one is left with a standard Ising model, where the energy of each spin configuration is replaced by its vibrational free energy. The efficient evaluation of the vibrational free energy of each configuration is the main problem limiting the inclusion of lattice vibrations in phase diagram calculations. A number of investigations have sought to assess the importance of vibrational effects on phase stability, in order to ensure that the efforts involved in computing vibrational properties are justified. The conclusion of the most reliable of these studies is that vibrational entropy differences are typically on the order of 0.1 k_B ot 0.2 k_B, which is comparable to the magnitude of configurational entropy differences (at most 0.69 k_B), thereby indicating that vibrations have a nonnegligible impact.

The calculation of the vibrational free energy of any configuration of the alloy reduces to the well known phonon problem in crystals. While the standard harmonic treatment of this problem lacks the ability to model thermal expansion, which can have a significant impact on thermodynamic properties in alloys, this limitation is easily overcome with the help of the quasiharmonic model. An exact solution to the phonon problem for all possible configurations requires excessive computing power. However, the tradeoff between accuracy and computational requirements can be controlled in two ways, namely through the selection of the range of force constants in the Born-von Kármán model, and through a choice of the number of ECI included in the cluster expansion describing the energetics of the Ising model. While there is evidence that the range of force constants can be kept very small (first nearest neighbor springs), the configurational dependence of the vibrational free energy is too complex to permit a drastic reduction in the number of ECI.

The main mechanisms that explain the trends in vibrational entropy differences between phases that have been suggested in the literature can be conveniently summarized with the "bond stiffness vs. bond length" model. In this picture, each type of chemical bond is characterized by a length-dependent spring constant. Changes in

vibrational entropy can originate from both changes in the proportion of each chemical bond and changes in their lengths as a result of local and global relaxations. This model not only provides an intuitive understanding of lattice vibrations in alloys, but also a practical way of including their effects in face diagram calculations. This stiffness vs. length relationship of each type of chemical bond can be inferred from a small number of lattice dynamics calculations. The vibrational properties of any configuration can then be obtained at a very low computational cost from the knowledge of the equilibrium geometry of this configuration, an information that is already a natural by-product of any phase diagram calculation.

There have so far been very few accurate phase diagram calculations that include the effect of lattice vibrations.The main limitation remains the determination of a cluster expansion that accurately models the configurational dependence of vibrational free energy. The "bond length vs. bond stiffness" model should prove to be an extremely useful tool in achieving this goal. Although this approximation has been very successful in all systems to which it has been applied, the confirmation of its validity in a wider range of systems is crucial.

Chapter 3

First-principles calculations of vibrational entropy

3.1 Introduction

Prior to this thesis, two mechanisms were commonly invoked to explain the configurational-dependence of vibrational entropy, namely, the "bond-proportion" and the volume effect. In this chapter, we intend to show that these two pictures are insufficient to correctly describe the coupling between an alloy's state of order and its vibrational entropy. This demonstration consists in performing accurate *ab initio* calculations of the vibrational change upon disordering in specific alloy systems and in identifying the source of the vibrational change through carefully designed thought experiments. The alloys systems chosen, the Pd-V and the Ni-Al systems, offer a unique opportunity to check the applicability of the "bond-proportion" and the volume mechanisms. Each of the two systems is expected to be a nearly ideal example of one of the two mechanisms.

The Ni-Al system is expected to exhibit a strong volume effect because of the large size mismatch (13%[1]) between the two species. In fact, Ni-Al is the system for which the volume mechanism was first introduced [2]. This importance of the volume effect was also found in subsequent semi-empirical calculations [8, 115] on the Ni-Al system.

In contrast, the Pd-V system appears an ideal "bond-proportion" system, for two reasons. First, the size mismatch between Pd and V is small (4%[2]) Second, Pd-V bonds are expected to exhibit a stiffness that exceeds the average stiffness of Pd-Pd and V-V bonds. The latter can be deduced from the characteristic trend of the bulk modulus of transition metals across the periodic table: Early transition metals are soft because their bonding orbitals are only partially filled. Stiffness reaches a maximum when the bonding orbitals are full and then decreases as the anti-bonding orbitals fill up. Pd and V are at each end of the transition metal section of the

[1]Determined by comparing the experimental lattice parameters of fcc Ni and fcc Al.

[2]This number was obtained by comparing first-principles calculations of the lattice parameters of fcc Pd and of V, artificially constrained to be in an fcc structure as well.

periodic table and are thus rather soft. When they are alloyed, the average number of valence electrons per atom is such that the bonding orbitals are closer to their optimal filling, suggesting that the stiffness of a Pd-V bond should be larger than the *average* stiffness of Pd-Pd and V-V bonds. (Note that since pure Pd is significantly stiffer than pure V, a Pd-V bond can still be softer than a Pd-Pd bond.)

The remainder of this chapter is organized as follows. We first describe the techniques used before discussing, in turn, the Pd-V and the Ni-Al systems.

3.2 Methodology

3.2.1 Theory

As described in the previous Chapter, in the harmonic approximation, the vibrational entropy of a structure can be obtained from the phonon density of states [81] $g(\nu)$. Above the Debye temperature of the solid, the high temperature limit is quickly reached and this dependence reduces to:

$$S_{vib} = -3k_B \left(1 + \ln\left(k_B T\right)\right) - k_B \int_0^\infty \ln\left(\nu\right) g(\nu) d\nu. \tag{3.1}$$

Since the first term is structure-independent, it has no effect on phase stability and will be ignored in the following analysis. The high-temperature limit is a good indicator of how large the effect of vibrations is likely to be in a given system, as vibrational entropy reaches its maximum in the high temperature limit. All vibrational entropies reported in this Chapter are calculated in the high-temperature limit.

The harmonic approximation can be made more realistic by allowing the phonon frequencies to be volume-dependent. This approach, called the quasi-harmonic approximation [81], enables the calculation of thermal expansion as well as its impact on the vibrational entropy. Once the volume-dependence of energy, $E(V)$, and vibrational entropy, $S_{vib}(V)$, are known, the equilibrium volume at temperature T is found by minimizing the free energy $F = E(V) - TS_{vib}(V)$ with respect to V. This technique has been used in previous computational investigations of the vibrational entropy [8, 106]. The accuracy of the quasi-harmonic approximation has been investigated in [46, 89]. Its accuracy was found to be excellent at temperature up to about the Debye temperature. At higher temperatures, its accuracy decreases and typically leads to a systematic overestimation of the thermal expansion. In this context, our results based on the quasi-harmonic approximation should be interpreted as an upper bound on the effect of thermal expansion.

Note that when thermal expansion is accounted for, the quantity $-k_B \int_0^\infty \ln\left(\nu\right) g(\nu) d\nu$ in Equation (3.1) becomes temperature-dependent, due to the change in volume. In the two systems considered here, this temperature-dependence is found to be small. For this reason, we will usually report high-temperature limiting values as a single temperature-independent quantity. However, one should not *a priori* expect this to be the case in all systems (see [8, 115, 106]).

3.2.2 *Ab Initio* calculations

The phonon density of states of an ordered compound can be accurately calculated through a variety of first-principles methods. Either the linear response technique [57, 40] or the fitting of a Born-von Kármán spring model to forces obtained from *ab initio* calculations [55, 143] can be used. The large computational requirements of the linear response technique limit its use to very symmetric small-cell structures. Since we will need to study structures with relatively large supercells and low symmetry, we rely instead on the fitting of spring constants to *ab initio* calculations of the forces acting on the atoms when they are perturbed away from their equilibrium positions.

Our *ab initio* calculations are performed within the local density approximation (LDA) using the VASP [75, 74] package which implements ultra-soft [139] pseudopotentials [113]. The precision of the quantum mechanical calculations used to determine the force constants is controlled by a variety of parameters. In the software package used, the electronic wavefunctions are represented by a linear combination of plane waves. As the number of planes waves is increased, the precision of the calculations increase because the wavefunctions can be represented more accurately. A convenient way to describe the number of plane waves used is to specify the so-called "energy cutoff", which is simply the kinetic energy of the plane waves having the shortest wavelength. There are essentially two types of error associated with the use of an insufficient number of plane waves First, the calculated energies and forces will obviously be contaminated by a random noise. Second, there is the problem of the "Pulay stress", which is a systematic bias of the calculated stress acting on the unit cell. Since the Pulay stress results in systematic errors in the calculated stress, the calculated equilibrium volume of the crystal will be biased as well. Since phonon frequencies are very sensitive to volume, this problem can introduce a systematic bias in the calculated vibrational properties. For this reason, a high cutoff was used when determining the equilibrium geometry of the structures (see Table 3.1). But once the equilibrium geometry is known, the cutoff can be safely reduced when computing the reaction forces acting on all atoms as one atom is moved away from its equilibrium position. The bias in the stress has essentially no impact on the forces acting on each atom. This technique achieves substantial savings in computer time since, for each structure, the equilibrium geometry has to be determined only once, while many force calculations are needed in order to determine all the force constants.

The second important parameter controlling the precision of the calculations, is the number of k-points used to sample the Brillouin zone. This parameter controls how accurately the band structure of the material is represented. It is important to note that as a structure's unit cell increases in size, the Brillouin zone is correspondingly reduced so that less k-points are needed to accurately represent the band structure. A convenient way to report the number of k-points used in a set of similar calculations with different unit cells is to specify the number of k-point by reciprocal atom, as done in Table (3.1).

How are the energy cutoffs and the number of k-points determined? The reaction forces in a simple test structure are calculated. Each parameter controlling the precision is gradually increased until it has essentially no effect on the calculated forces,

Parameter	Ni-Al	Pd-V
Energy cutoff for relaxations	400 eV	365 eV
Energy cutoff for force calculations	241 eV	211 eV
Number of k-points	$(14)^3/(atom^{-1})$	$(14)^3/(atom^{-1})$

Table 3.1: Parameters Used in the First-Principles Calculations.

yielding what can be considered as an "exact" result. By comparison with this "exact" result, one can see for which values of the cutoff and k-point parameters the precision of the calculated forces is sufficient to obtain the desired accuracy in the vibrational entropy. In order to map the desired level of precision in the vibrational entropy onto a required precision in the forces, we use the following rule.[3]

$$\frac{|\delta S_{vib}|}{N} \leq \frac{3}{2} k_B \max_j \left| \frac{\delta f_j}{f_j} \right| .$$

where $\max_j \left| \frac{\delta f_j}{f_j} \right|$ is the maximum fractional error on the calculated forces.

In our calculations, we have required the forces to differ by no more than 0.5% from the "exact" value, which should be sufficient to obtain vibrational entropies with a precision of the order of $0.007 k_B$.

3.2.3 The disordered state

While the calculation of the vibrational properties of the ordered state is straightforward, the case of a disordered alloy presents numerous difficulties associated with large computational requirements. The most direct way to model the disordered state

[3]This rule can be derived as follows. In the high-temperature limit,

$$\frac{S_{vib}}{N} = \frac{k_B}{N} \sum_{i=1}^{3N} \ln \sqrt{\lambda_i} + \text{constant}$$

where the λ_i are the eigenvalues of the $3N \times 3N$ dynamical matrix of the system. We can the write

$$\begin{aligned} \frac{\delta S_{vib}}{N} &= \frac{1}{2} \frac{k_B}{N} \sum_{i=1}^{3N} \left(\frac{\delta \lambda_i}{\lambda_i} \right) \\ &= \frac{3}{2} k_B \left\langle \frac{\delta \lambda_i}{\lambda_i} \right\rangle . \end{aligned}$$

If all the calculated forces f_j were shifted by the same fraction $\frac{\delta f}{f}$, the resulting eigenvalues would also be shifted by the same fraction. While the errors in the forces tend to random rather than systematic, assuming them to be systematic provides a convenient upper bound on the errors:

$$\frac{|\delta S_{vib}|}{N} \leq \frac{3}{2} k_B \max_j \left| \frac{\delta f_j}{f_j} \right| .$$

Structure	p_1	p_2	t_{111}	t_{112}	t_{113}	t_{114}
$L1_2$	0	1	1/2	-1/2	1/2	-1/2
$D0_{22}$	0	2/3	1/2	-1/6	1/6	1/6
$\|L1_2 - DO_{22}\|$	0	1/3	0	1/3	1/3	2/3
SQS-8	1/4	1/3	-1/4	0	-1/12	-1/6
Random	1/4	1/4	-1/8	-1/8	-1/8	-1/8
|SQS-8 − Random|	0	1/12	1/8	1/8	1/24	1/24

Table 3.2: Correlations of the Structures Used. p_n denotes the n-th nearest neighbor correlation while t_{lmn} denotes a triplet made of overlapping p_l, p_m, and p_n pairs.

is to rely on a large supercell calculation where the occupation of the lattice sites is randomly chosen. Unfortunately, both the linear response and the spring constant fitting approaches become impractical for very large supercells.

A computationally efficient way to model the disordered state is to rely on a so-called Special Quasirandom Structure [151] (SQS). A SQS is the periodic structure that best approximates the disordered state in a unit cell of a given size. The SQS approach has been used very successfully to obtain electronic and thermodynamic properties of disordered materials (see, for example, [63]). The accuracy of the SQS approach has been benchmarked using the embedded atoms potentials, which allow the computation of the vibrational entropy of a large supercell simulating the disordered state.[90] An SQS having only 8 atoms in its unit cell was found to already provide a good approximation of the disordered state for the purpose of calculating vibrational properties.

The quality of a SQS is described by the range within which the correlations ρ_α between the occupation of different lattice sites mimics the ones of the disordered state. For an fcc lattice at concentration 3/4, the 8-atom SQS shown in Figure 3-1a) is able to reproduce the nearest-neighbor pair correlation of the disordered state exactly. Other longer-range and multi-body statistical correlations are approximately reproduced, as shown in Table 3.2. The magnitude of the errors introduced by the approximation of these other correlations can be estimated by computing the vibrational entropy of different structures which have the same nearest-neighbor correlation, but different longer-range and multi-body correlations. On the fcc lattice at concentration 3/4, the two simplest structures that have this property are the $L1_2$ and the DO_{22} structures. If the vibrational entropy of the $L1_2$ and the DO_{22} structures differ by an amount $\Delta S^{L_{12}\rightarrow DO_{22}}$, the difference between the 8 atom SQS and the true disordered state is expected to be half as much, since their longer-range and multi-body correlations are more similar, as shown in Table 3.2. Table 3.3 reports $\Delta S_{vib}^{L_{12}\rightarrow DO_{22}}$ in both the Ni-Al and Pd-V systems, along with the corresponding estimate of the error introduced by the SQS approximation.

3.2.4 Force constants calculations

The force constants of a Born-von Kármán model are determined through the procedure described in Section 2.4.1. Each distinct atom is, in turn, displaced away from

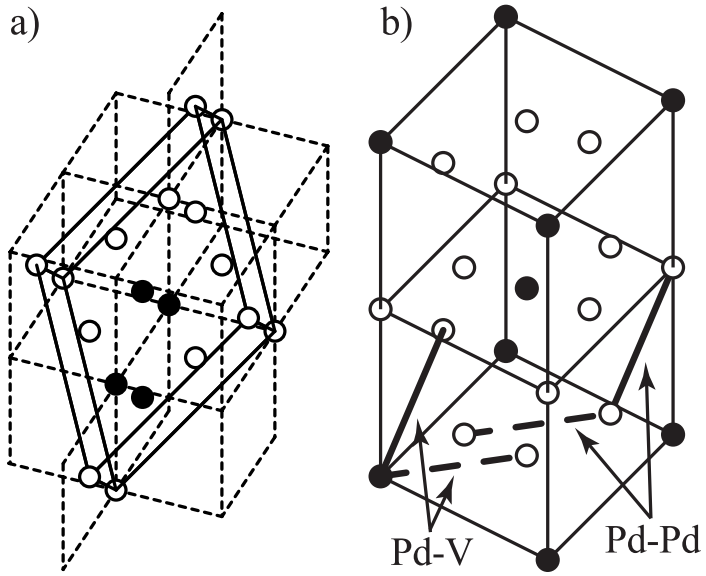


Figure 3-1: a) 8-Atom SQS Used to Model the Disordered State. (Primitive unit cell shown.) b) Constraints on Bond Lengths Originating from the Symmetry of the Ordered $D0_{22}$ structure. (Conventional cell shown.) Bonds represented by identical line styles have identical lengths.

System	$\Delta S_{vib}^{L_{12}\rightarrow DO_{22}}$	$\Delta S_{vib}^{SQS-8\rightarrow\text{random}}$
Ni-Al	$0.04k_B$	$\pm 0.02k_B$
Pd-V	$0.08k_B$	$\pm 0.04k_B$

Table 3.3: Differences in Vibrational Entropy Between the $L1_2$ and the $D0_{22}$ Structures in the Ni-Al and Pd-V Systems. Also shown is the implied precision of the SQS approximation.

its equilibrium position and resulting forces on all atoms are calculated from first principles. This information can then be used in a least square fitting procedure to find the spring constants that best predicts the forces acting on each atom. This procedure is controlled by numerous parameters that need to be chosen to maximize the precision of the calculated force constants at a reasonable computational cost.

The first parameter is the magnitude of the atomic displacements. If the displacements are chosen too small, the inaccuracies in the calculated forces will be large relative to the forces themselves and the overall precision of the procedure will decrease. If the displacements are chosen too large, the anharmonicity of the potential surface will affect the precision of the harmonic force constants. In order to find the best balance between these two effects, we need two pieces of information: the precision of the forces Δf and the magnitude of the anharmonicity. While the precision of the forces Δf is known (from the study of the convergence with respect to the energy cutoff and the number of k-points performed in Section 3.2.2), an estimate of the anharmonicity can be obtained from the following analysis. We select one type of atomic displacement as a test case and try different magnitude of displacements x. We then fit the resulting forces f to a polynomial:

$$f = a_2 x + a_3 x^2 + a_4 x^3.$$

Since the third order term $a_3 x^2$ is easily be cancelled by considering positive and negative displacements, we focus on the error introduced by the fourth order term $a_4 x^3$. The precision of the vibrational entropy depends on the *relative* error on the forces. Therefore, the optimal displacement can be found by minimizing

$$\frac{\Delta f + a_4 x^3}{a_2 x}$$

with respect to x. In both the Pd-V and Ni-Al systems, the optimal displacement was found to be about 0.05 Å.

Another important parameter controlling the precision of the spring model is the range of interactions considered. This parameter can be determined by gradually including longer-range spring interactions until the value of vibrational entropy converges. The study of the convergence of the vibrational entropy as a function of the range of springs considered, shown in Table 3.4, reveals that, in the Ni-Al system, even a first nearest neighbor spring model already gives us an accuracy of $0.025 k_B$. In the Pd-V system, the convergence of the absolute value of vibrational entropy is much slower (see Table 3.5). Fortunately, we are interested in vibrational entropy differences, whose convergence is much quicker. Table 3.5 shows that a nearest neighbor spring model is able to provide vibrational entropy differences with an accuracy of about $0.02 k_B$. In the event that this rapid convergence of vibrational entropy differences is fortuitous and unique to the $L1_2$ and $D0_{22}$ structures, the speed of convergence of the absolute vibrational entropies, rather than their differences, should be used as a measure of precision. Although this pessimistic estimate is of the order of $0.1 k_B$, our results remain conclusive in the presence of an error of this magnitude.

Structure	1st NN Shell	2nd NN Shell	3rd NN Shell
$L1_2$	-5.550	-5.576	**-5.575**
SQS-8	**-5.569**		

Table 3.4: Vibrational Entropy (in k_B) as a Function of the Number of Nearest Neighbor (NN) Shell Included in the Born-von Kármán Model.

Structure	1	2	3
$L1_2$	-4.39	-4.44	-4.48
$D0_{22}$	**-4.47**	-4.53	-4.58
SQS-8	**-4.54**		
$L1_2-D0_{22}$	0.08	0.08	0.10
SQS-8$-D0_{22}$	-0.07		

Table 3.5: Vibrational Entropy (in k_B) as a Function of the Interaction Range Included in the Spring Model. Range is expressed as the number of nearest neighbor shells.

The high accuracy of such a short-range spring model is not unusual: It has been observed [55] that even though a long-range spring model is required to model all the features of the phonon DOS, an integrated quantity such as the vibrational entropy converges much faster with respect to the range of interaction included. We exploit this fact to describe the disordered state, where longer ranged interactions would be prohibitive to calculate, using nearest-neighbor spring constants only. Note that due to the low symmetry of the SQS, a nearest-neighbor model still involves the evaluation of 50 distinct parameters in the spring tensors.

Now that the computational techniques have been described, we are ready to interpret the results of these calculations.

3.3 The Pd-V system

3.3.1 Results

Our main result is that the calculated vibrational entropy of the disordered state is $0.07k_B$ *lower* than the one of the ordered state ($D0_{22}$ structure), as seen by comparing the two numbers in bold in Table 3.5. This is in complete contradiction with the well-known "bond proportion" mechanism which would predict a large and *positive* change in vibrational entropy.

Before we discuss the origin of this unexpected result, a few remarks are in order. Since the convergence rate of the vibrational entropy with respect to the range of interaction is relatively slow in Pd_3V, we compare the vibrational entropy between two spring models having the same range of interactions, in order to ensure the maximum amount of cancellation of any potential systematic errors. It is also important to note that even using the most pessimistic estimate of the precision of our approach ($0.1k_B$), the vibrational entropy change upon disordering is no larger than $0.03k_B$,

Quantity	Units	$D0_{22}$	SQS-8
Bulk Modulus	GPa	215	191
Atomic Volume	$Å^3$/atom	13.808	13.891
Average Grüneisen parameter		2.41	2.24
Linear Thermal Expansion Coef.	$10^{-6}K^{-1}$	11.2	11.6
Temperature-Dependence of S_{vib}	$10^{-6}k_B$/(K atom)	243	234

Table 3.6: Calculated Properties of the Ordered ($D0_{22}$) and the Disordered State (Approximated by an 8 Atom SQS). Bulk modulus and volume are given at 0K while the other quantities are the high-temperature limiting values.

which is small compared to typical vibrational entropy changes, which are of the order of $0.1k_B$.

A quasi-harmonic treatment enables the evaluation of the temperature-dependence of this entropy difference. While the temperature dependence of the vibrational entropy is large in both the ordered and the disordered state (see Table 3.6), they are almost identical and have little impact on phase stability. This identity is remarkable: the ordered and disordered materials have a different bulk modulus and a different Grüneisen parameter (see Table 3.6), but these two differences offset one another.

Another unexpected finding is that the bulk modulus of the disordered state is smaller than that of the ordered state, which is in apparent contradiction with our claim that the disordered state is "stiffer" than the ordered state. However, one must keep in mind that vibrational entropy depends on the average stiffness of all possible vibrational modes, while bulk modulus measures the stiffness of only one mode. The Pd-V system thus provides an example where estimates of the vibrational entropy based on bulk modulus, such as the Debye-Grüneisen model [92], can be misleading.

3.3.2 Discussion

The fact that the vibrational entropy change upon disordering does not have the sign that one would expect for an ordering system merits further analysis. This section demonstrates that the origin of this surprising result can be traced back to the effect of local relaxations. We first present a simple model that allows us to isolate the origin of the vibrational entropy differences in this system. We then introduce a precise mechanism that is able to explain our results before proceeding to show that this particular mechanism is indeed at work in Pd_3V.

Consider a simplified spring model obtained by including only stretching and bending terms in the first nearest-neighbor spring tensors. These terms can be read from the diagonal elements of the spring tensor associated with a given pair of atoms, when this tensor is represented in a cartesian basis with one axis aligned along the segment joining the two atoms in question. In addition, the bending terms are constrained to be orientation-independent. The resulting tensor contains only two independent

Structure	1(len)	1(sb)	1
$L1_2$	-4.39	-4.40	-4.39
$D0_{22}$	-4.42	-4.48	-4.47
SQS-8	-4.56	-4.53	-4.54
$L1_2-D0_{22}$	0.03	0.08	0.08
SQS-8$-D0_{22}$	-0.14	-0.05	-0.07

Table 3.7: Vibrational Entropy (in k_B) Obtained with Various First Nearest Neighbor Spring Models. The full spring tensors is used in Column 1. Only stretching and bending terms are included for the column labeled 1(sb) while the column labeled 1(len) presents the results of a model where bond stiffness is only allowed to depend on bond length only (see text).

spring constants, a stretching term s and a bending term b:

$$\begin{pmatrix} -s & 0 & 0 \\ 0 & -b & 0 \\ 0 & 0 & -b \end{pmatrix}.$$

The error introduced by this approximation (relative to using the full first nearest-neighbor tensor) never exceeds $0.01k_B$ for all fcc structures tested (see Table 3.7, columns 1(sb) and 1.). In these calculations, the simplified spring model is applied to the fully relaxed geometries, as determined from *ab-initio* calculations. Note that keeping only stretching terms would be an oversimplification, as it can result in errors comparable in magnitude to the typical values of the vibrational entropy of formation ($\pm 0.2k_B$). These errors are non-systematic and do not cancel out when taking entropy differences.

This simple spring model is useful from a conceptual point of view, as it lets us compare the stiffness a given type of bonds (e.g. Pd-Pd, V-V or Pd-V) in different structures: the spring tensors have the same form regardless of the symmetry of the bonds' environment. Figure 3-2 shows the values of the stretching (s) and bending (b) terms of the spring tensor of bonds of various lengths taken from a set of fcc-based structures ($L1_2$, $D0_{22}$, SQS-8, fcc Pd and fcc V, each taken at two different volumes). Bond stiffness correlates reasonably well with bond length, as seen by the least squares fit shown in Figure 3-2. Bond stiffness typically decreases with bond length. The fact that the same relationship between bond stiffness and length holds throughout different structures is a important feature that will leads us to a simple mechanism explaining our results for the Pd-V system. While both the stretching and bending terms are important to consider for quantitative purposes, the magnitude of the stretching term only provides a convenient measure of a bond's stiffness for the purpose of the following qualitative discussion.

We argued earlier that a Pd-V bond resembles a bond between two elements of the middle of the transition metal series, which typically have a larger stiffness. However, elements in the middle of the transition metal series are also characterized by smaller lattice constants. One would then expect Pd-V bonds to be shorter than the average

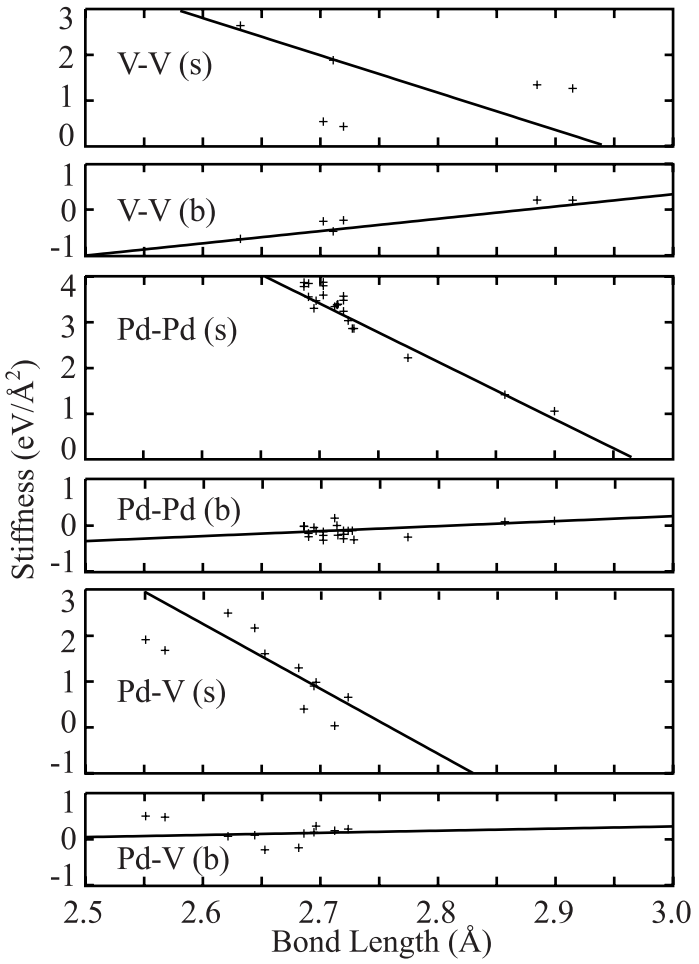


Figure 3-2: Stretching (s) and Bending (b) Terms of the Nearest-Neighbor Spring Tensor as Function of Bond Length. Each point corresponds to one type of bond in one of a set of fcc structures ($L1_2$, $D0_{22}$, SQS-8, fcc Pd and fcc V, each taken at two different volumes).

Bond	Length (Å)	Stiffness (eV/Å^2)
Pd-Pd	2.743	3.06
V-V	2.763	0.69
Average	2.753	1.45
Pd-V	2.628	2.21

Table 3.8: Average Bond Length and Bond Stiffness (Along the Stretching Direction) in the Disordered State. The row labeled "average" reports the arithmetic average of the length of Pd-Pd and V-V bonds and the geometric average of their stiffness

Bond	Length (Å)	Stiffness (eV/Å^2)
V-Pd	2.693	0.61
Pd-Pd	2.693	3.44

Table 3.9: Average Bond Length and Bond Stiffness (Along the Stretching Direction) in the Ordered State.

length of V-V and Pd-Pd bonds. As shown in Table 3.8, the average bond length and stiffness in the disordered state are in perfect agreement with this picture. Note that, while Pd-V bonds are stiffer than the *average* stiffness of Pd-Pd and V-V bonds, as expected, Pd-V bonds are nevertheless softer than Pd-Pd bonds.

The situation is quite different in the ordered state: the high symmetry of the structure constrains the Pd-V bonds to have the same length as the Pd-Pd bonds (see Figure 3-1b)). The average bond length tends to be much closer to the Pd-Pd "ideal" length than to the Pd-V "ideal" length because Pd-Pd bonds are stiffer than Pd-V bonds. The result is an ordered alloy where Pd-V bonds are significantly longer than they would be in the absence of symmetry constraints while the Pd-Pd bond lengths are only slightly affected. Pd-V bonds are therefore unusually soft in the ordered state, while the stiffness of Pd-Pd bonds is nearly unaffected. This tends to makes the ordered state softer and is responsible for its higher vibrational entropy. The average bond length and stiffness in the ordered state shown in Table 3.9 support this interpretation.

The fact that disordering shortens the Pd-V bond while leaving the Pd-Pd bonds mostly unchanged on average can be seen from the histogram of the bond length distribution (Figure 3-3). The impact of these bond length changes on stiffness is best illustrated by plotting the change in average bond length and stiffness upon disordering, as illustrated in Figure 3-4. The dramatic stiffening of the Pd-V bonds and the slight softening of the Pd-Pd bonds in the disordered state, relative to the ordered state, is clearly visible.

Static displacements of this magnitude should be visible in diffuse scattering measurements. Such measurements have been performed in Pd_3V [124] and in a related system, Pt_3V [22]. One of the authors of reference [124] (F. Ducastelle) has indicated to us that the more precise measurements made on Pt_3V should give us a reliable upper bound on the magnitude of the static displacements in disordered Pd_3V, where the determination of the static displacements was less precise. We will thus compare

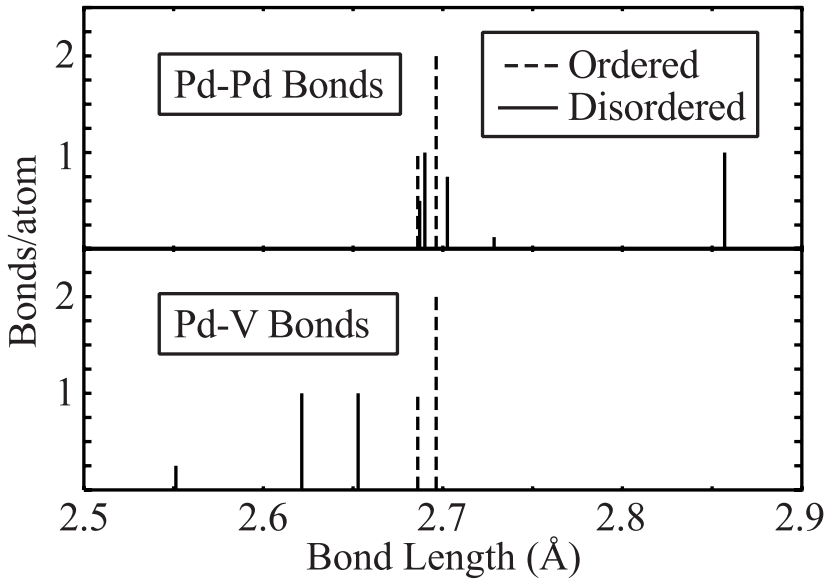


Figure 3-3: Bond Length Distribution.

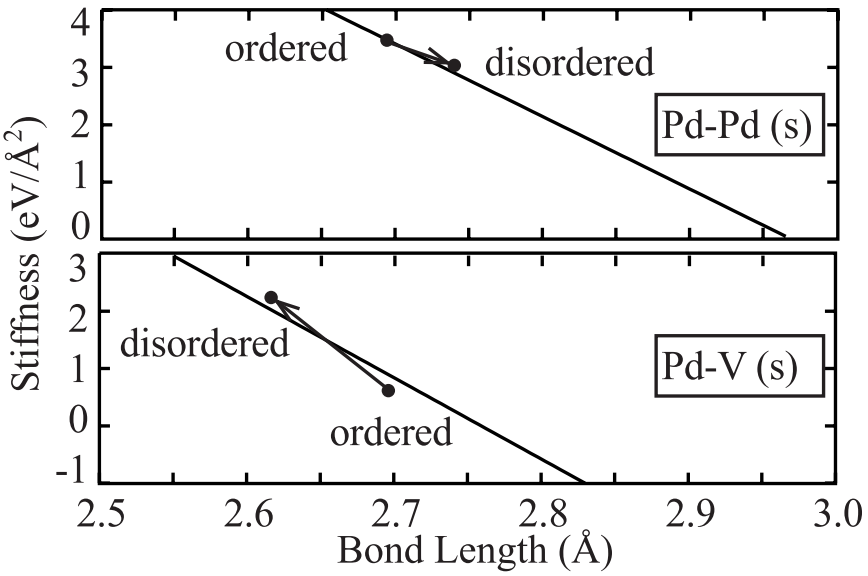


Figure 3-4: Shift in Average Bond Stiffness (Along the Stretching Direction) and Bond Length upon Disordering. The fitted line of figure Figure 3-2 is shown for reference.

our results with the Pt_3V measurements only.

As the scattering factor of V is much smaller than the one of Pt, it is difficult to measure shifts in the Pt-V and V-V bond lengths. Unfortunately, these are precisely the bonds we predict to be the most affected by disordering. The experimental nearest neighbor average Pt-Pt bond length shift is reported to be 0.3% of the lattice parameter, which is somewhat smaller than ours (1.2% of the lattice parameter). This discrepancy can be easily explained by the fact that we model disordered Pd_3V as a perfectly disordered material, while disordered Pd_3V actually exhibits short-range order. Fully disordered Pd_3V is naturally expected to exhibit larger relaxations. Note that the presence of short-range order does not invalidate our discussion. In the presence of short-range order, both the traditional 'bond proportion' mechanism and the effect of relaxations will decrease in importance, but they would still give rise to competing and comparable contributions to the vibrational entropy change, which is our main observation. As such, existing experimental observations do not contradict our findings. Unfortunately, the most salient feature of our predicted static displacements, the shortening of Pd-V bonds, has not yet been confirmed experimentally.

Perhaps the easiest way to separate the effect of the "bond proportion" mechanism from the effect of relaxations is to construct a model system where bonds always have the opportunity to reach their "ideal" length, regardless of the symmetry of their local environment. The average stretching and bending force constants obtained in the disordered state, listed in Table 3.8, are used as an approximation to the "true" force constants that would be expected in the absence of symmetry constraints. These force constants are used to calculate the vibrational entropy for both the ordered DO_{22} and SQS-8 structures. The vibrational entropy change upon disordering then becomes $0.26k_B$, which is large and positive, as expected when the "bond proportion" mechanism operates alone. The large configurational dependence of vibrational entropy provided by the "bond proportion" mechanism is thus entirely masked by relaxation effects to yield vibrational entropy difference of $-0.07k_B$.

While the above model system is useful for illustrative purposes, we have to verify that the difference between $0.26k_B$ and $-0.07k_B$ can really entirely be attributed to the effect of relaxations. For instance, this difference includes the error introduced by replacing each bond's force constants by average force constants. Vibrational entropy is not a linear function of the force constants, and averaging the latter could bias the former. Moreover, bond stiffness could vary for reasons other than bond length change: for example, the local charge density in the neighborhood of a given bond could vary. For these reasons, we now introduce a model system which (1) does not rely on averaged force constants and (2) only accounts for bond stiffness change due to bond length changes.

To show that the effect of relaxations alone can explain our results, we make use of the *"bond stiffness vs. bond length" model*. We replace the true stiffness of each bond by the one predicted from bond length through a simple least squares fit (shown in Figure 3-2). While this simplified model exhibits a limited accuracy (see column 1(len) of Table 3.7), it is clearly able to predict that the vibrational entropy of the disordered state is lower than the one of the ordered state. In the simplified model, a bond's stiffness is uniquely determined by its type (Pd-Pd, Pd-V or V-V) and its

length. Variations in bond stiffness that are not due to bond length are ignored, leaving only relaxations as the possible source of the higher stiffness of the disordered state.

It is worth noting that our suggestion of defining bond characteristics that are transferable (*i.e.* applicable to different structures) bears some resemblance to an earlier attempt to define transferable "configuration averaged force constants" (CA FC) [123]. However, our approach differs in three important respects. First, we keep only stretching and bending terms in the spring tensors, thus avoiding the incompatibilities in the form of the spring tensor when the symmetry environment of a bond differs in distinct structures. Second, we do not try to define a universal bond-specific stiffness but instead define a universal stiffness versus length relationship. The stiffness of a bond is thus allowed to vary in different structures when its length varies. Finally, we do not attempt to define force constants that also predict the correct equilibrium geometry of a structure. In a typical *ab- initio* phase diagram calculation, the exact equilibrium geometry is already known, as it is a byproduct of the calculation of the energy of a given structure.

3.3.3 Conclusion

The ordering tendency of the Pd-V system would indicate that Pd-V bonds should be stiff relative to Pd-Pd and V-V bonds. Based on this observation, one would expect the vibrational entropy change upon disordering to be large and positive. Instead, our calculations indicate a negative vibrational entropy change. The origin of this surprising result lies in the fact that the Pd-V bonds are stiffer only when they are allowed to relax to their short "ideal" length, which can happen in the disordered state but not in the ordered state, due to symmetry constraints. The stiffening of the Pd-V bonds in the disordered state more than compensates for the fact that there are less Pd-V bonds in the disordered state. The larger stiffness of the disordered state translates into a vibrational entropy that is lower than the one of the ordered state. We prove that this mechanism indeed determines the observed sign of the vibrational entropy change upon disordering through two model systems. In one system, we entirely remove the effect of relaxations and find that the vibrational entropy change now has the positive sign typically expected in a ordering system. In a second model system, we only include bond stiffness changes that can be associated with bond length changes and find that the vibrational entropy decreases upon disordering, in agreement with the results of our more accurate calculations. These two results unambiguously show that relaxations play an essential role in determining vibrational entropy changes in the Pd-V system. These results also illustrate both the inadequacies of the "bond proportion" picture and the predictive power of the "bond stiffness vs. bond length" model.

3.4 The Ni-Al system

3.4.1 Background

The Ni_3Al compound is of "historical" importance in the study of lattice vibration in alloys. It was the first system where the vibrational entropy change associated with a disordering reaction was directly measured experimentally. It was also the first direct experimental evidence that the thermodynamic effect of vibrational excitations can be comparable in magnitude to the effect of configurational excitations: the vibrational entropy difference was estimated to be about $0.2k_B$ [49, 11] whereas the configurational entropy difference can be at most $0.57k_B$. This finding attracted the attention of alloy theorists and many theoretical calculations of the vibrational entropy of disordering of the Ni_3Al compound followed [2, 8, 115]. To keep computational requirements in control, these calculations of the vibrational entropy difference between disordered and ordered Ni_3Al (hereafter denoted $\Delta S_{vib}^{o\rightarrow d}$) were performed using the semi-empirical Embedded Atom Method (EAM) [38]. Although the specific result seemed to depend somewhat on the EAM potential used, all authors found values between $0.1k_B$ and $0.3k_B$, which corresponded to the range of values found experimentally.

Although the value of $\Delta S_{vib}^{o\rightarrow d}$ calculated through EAM agrees with experimental results, there is no consensus on the origin of this difference. The EAM results indicate that disordering causes a nearly uniform softening of all phonon modes due to an increase in volume. On the other hand, it has been argued from neutron scattering experiments [49] that the vibrational entropy increase is mainly due to a significant decrease of the number of high frequency vibrational modes. Surprisingly, the samples used to measure vibrational entropy differences even exhibited a *decrease* in lattice constant upon disordering, in contrast to the EAM results.

Some EAM investigations [115] found that a large fraction of $\Delta S_{vib}^{o\rightarrow d}$ is due to thermal expansion differences between the ordered and the disordered state, while experiments [11, 49], as well as some theoretical investigations [2], obtained a large $\Delta S_{vib}^{o\rightarrow d}$ without considering this effect. Finally, recent EAM calculations [88] (which update some of the results found in [8]) found essentially no contribution from thermal expansion.

We will attempt to reconcile some of these conflicting observations through a first-principles calculation of the difference in vibrational entropy between the ordered and disordered state for the Ni_3Al compound. We will also identify the precise mechanism that determines the change in vibrational entropy upon disordering the Ni_3Al compound.

3.4.2 Results

Before addressing our main question, we assess the accuracy of our first principles calculations by comparing various experimentally determined quantities with the results of our calculations. Figure 3-5 compares the measured heat capacity of ordered Ni_3Al with the one calculated from first-principles. The agreement is very good below 1000

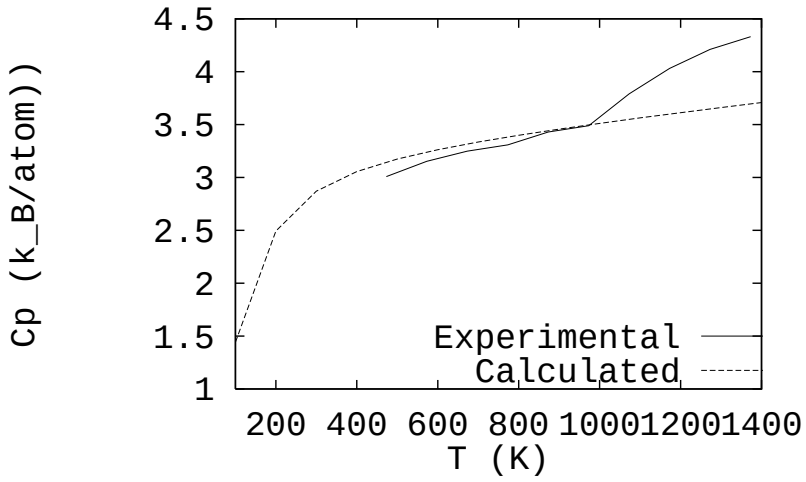


Figure 3-5: Calculated and Experimental Heat Capacities (at Constant Pressure) as a Function of Temperature.

K but apparently degrades quickly beyond that temperature. The reason for this apparent discrepancy is simply that configurational excitations, which are purposely ignored in our calculations, start to contribute significantly to the heat capacity at high temperature [8]. The agreement between the calculated and measured thermal expansion, shown in Figure 3-6, is even more striking.

Figure 3-7 shows how our calculated phonon DOS of the $L1_2$ structure compares with experimental results [49]. Due to the fact that the Local Density Approximation (LDA) systematically underestimates lattice constants, the calculated frequencies are slightly too high. However, since a similar effect is present for both the ordered and the disordered state, this bias is expected to mostly cancel out. A more satisfactory solution to this problem will be presented in Chapter 4.

Calculated and Experimental Phonon DOS of the $L1_2$ Structures. The experimental phonon DOS is calculated from the force constant provided in [49], which are themselves fitted from previous phonon dispersion measurement [126].

We now turn our attention to our main finding. The results shown in Table 3.4.reveal that there is almost no difference in vibrational entropy between the ordered and the disordered state ($0.006k_B$). This value is obtained by subtracting the vibrational entropy of the ordered state obtained with a third nearest neighbor spring model from the vibrational entropy of the disordered state obtained from a first nearest neighbor spring model. An error bracket of $\pm 0.05k_B$ can be attributed to this result by adding the uncertainties due to using only nearest neighbor springs for the SQS ($0.025k_B$) and due to using a SQS having only the correct nearest neighbor pair correlations ($0.02k_B$).

We do find a large temperature-dependence of S_{vib} for both the ordered and the disordered states (about $2.4 \times 10^{-4} k_B/\mathrm{K}$). However, since the *difference* in temperature-

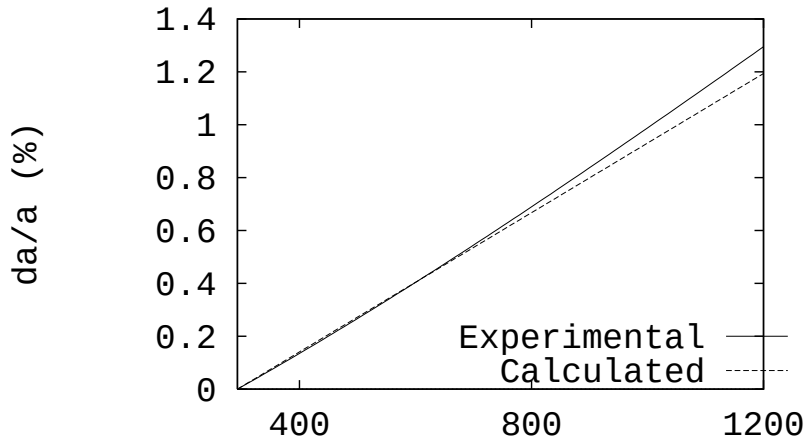


Figure 3-6: Calculated and Experimental Lattice Parameters as a Function of Temperature.

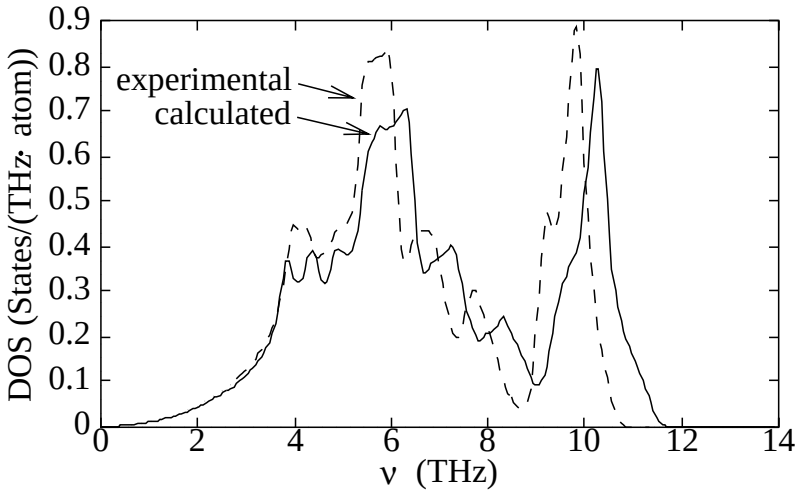


Figure 3-7: Calculated and Experimental Phonon DOS of the $L1_2$ Structures.

Quantity	Units	$L1_2$	SQS-8
Bulk Modulus	GPa	204	179
Atomic Volume	Å^3/atom	10.627	10.680
Average Grüneisen parameter		2.07	1.94
Linear Thermal Expansion Coef.	$10^{-6}\mathrm{K}^{-1}$	13.2	14.0
Temperature-Dependence of S_{vib}	$10^{-6}k_B$/(K atom)	244.7	245.3

Table 3.10: Calculated Properties of the Ordered ($L1_2$) and the Disordered State (Approximated by an 8 Atom SQS). Bulk modulus and volume are given at 0K while the other quantities are the high-temperature limiting values.

dependence between the two states is very small ($0 \pm 1 \times 10^{-5} k_B/\mathrm{K}$), $\Delta S_{vib}^{o \to d}$ remains small at all temperatures. This identity is remarkable: the ordered and disordered materials have a different bulk modulus and a different Grüneisen parameter (see Table 3.10), but these two differences precisely offset one another.

3.4.3 Discussion

How can we explain the apparent discrepancy between our findings ($0.00 \pm 0.05 k_B$) and results from inelastic neutron scattering [49] (from 0.1 to $0.3 k_B$)? Although this range of experimental results is often averaged to give $0.2 k_B$, the true result is probably closer to $0.1 k_B$. The higher bound of $0.3 k_B$ was obtained using the virtual crystal approximation to analyze the neutron scattering data of the disordered state, while the lower bound of $0.1 k_B$ was obtained when the DOS of the disordered state was assumed to resemble that of the ordered state. Recent EAM calculations [8, 115], as well as our own results from the SQS calculations (see Figure 3-8), clearly show the latter hypothesis to be the correct one: The DOS of the disordered state is a broadened version of the DOS of the ordered state. The experimental result is therefore likely to be close to the lower bound of $0.1 k_B$. The remaining discrepancy between our calculation and experiment can well be accounted for by other sources of entropy in the experiment.

As Ni_3Al is ordered up to the melting point, the metastable disordered phase of Ni_3Al needs to be produced by vapor deposition [61] or ball milling [66]. Both of these methods produce a material with a very small grain size (respectively 4 nm [61] and 7 nm [47]). When the samples are annealed to reach the ordered state, the grain size inevitably increases. It follows that the ordered and disordered samples differ not only by their state of order, but also by their grain size. Grain size can have a significant effect on the thermodynamics of nanocrystalline materials. The latter typically possess higher heat capacities than their coarse-grained counterparts due to the large fraction of atoms lying in the grain boundary regions which are typically softer [131, 47, 51]. For example, it has been observed that the vibrational entropy difference between disordered Ni_3Fe in its nanocrystalline (9 nm average grain size) and coarse-grained form is about $0.18 k_B$ [47], which is of the same order of magnitude as $\Delta S_{vib}^{o \to d}$ in Ni_3Al.

This small grain size effect is visible in the vibrational DOS obtained through neu-

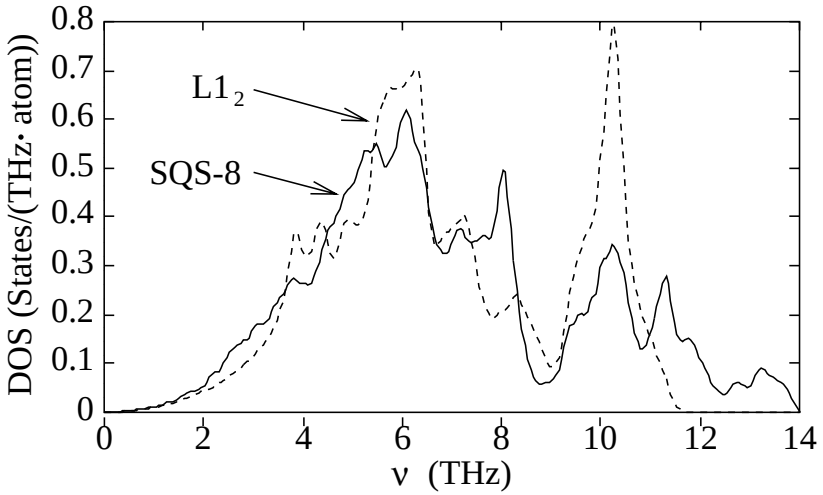


Figure 3-8: Calculated Phonon DOS of the $L1_2$ and SQS-8 Structures.

tron scattering. It is responsible for the enhancement of the density of low frequency phonon modes [49]. We obtain an estimate of this effect by integrating the experimentally measured $\ln(\nu)(g^o(\nu) - g^d(\nu))$ over the low frequency part of the DOS. To yield a meaningful value, this integral has to be taken over a range of frequencies which encloses the same number of modes for both the ordered and the disordered state. From the data of Fultz *et al.* ([49], Figure 4), we obtain a low frequency contribution of about $0.05k_B$ by integrating from 0 to 22 meV. Frequencies above 22 meV provide the remaining 0.05 k_B attributable to the order-disorder transition. With this new interpretation, the neutron scattering results now lie in between the EAM and the ab-initio predictions.

In calorimetric measurements [11], grain size effects are expected to be even larger, as the grain size was only 4 nm (about 20 atomic layers), compared to 7 nm in neutron scattering experiments. This partly explains the larger value of the vibrational entropy difference found with calorimetric measurements.

The effect of small grain size can also explain the disagreement between the calculated and experimental lattice parameters. All calculations indicate that the disordered state has a larger volume than the ordered state (1%–2% larger with EAM, 0.5% larger with ab-initio). On the contrary, some experiments [49] (but not all: See [26]) find that the ordered state has the largest lattice constant. The difference arises from the fact that the lattice constant of nanocrystalline materials can often differ by $\pm 0.3\%$ from their bulk counterparts [4, 149, 79]. This effect can dominate over the volume expansion of the bulk material only if the latter is not too large. In this sense, the small volume expansion upon disordering we obtain is more consistent with experimental observations than the EAM results.

In summary, the large $\Delta S_{vib}^{o\to d}$ found in experiments can, in fact, be attributed to a variety of parasitic effects that can very easily be avoided in *ab initio* calculations. The large $\Delta S_{vib}^{o\to d}$ found in semi-empirical calculations is due to an overestimation of the volume expansion upon disordering. The smaller volume expansion found in *ab initio* calculations is much more consistent with experimental observations.

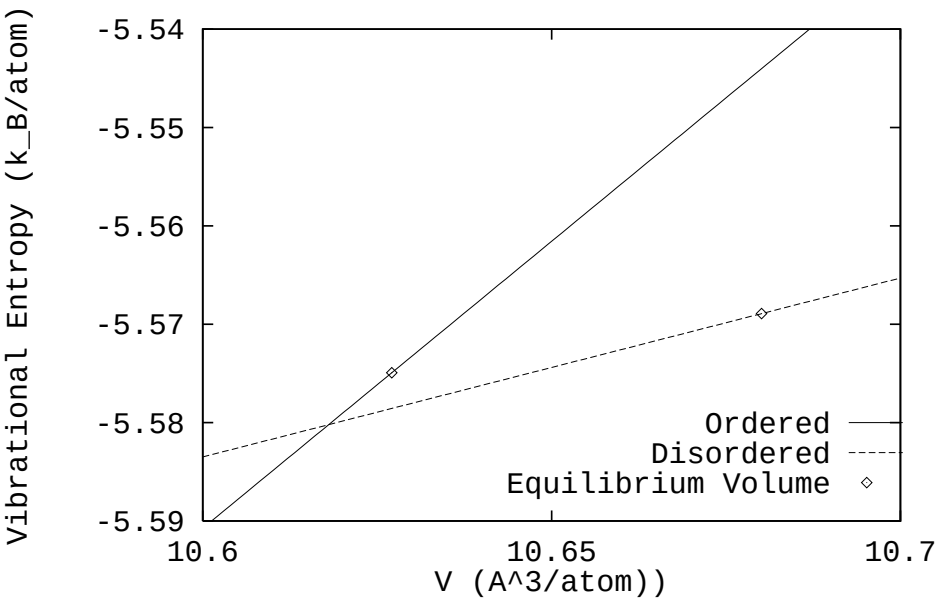


Figure 3-9: Vibrational Entropy as a Function of Atomic Volume in the Ordered and the Disordered States.

3.4.4 The origin of the small vibrational entropy change upon disordering

While accurate *ab initio* calculations were helpful in reconciling a number of conflicting observations we still need to understand the mechanisms that are responsible for this small vibrational entropy change upon disordering. First-principles calculations did predict a small increase in volume upon disordering and it is surprising that the associated increase in vibrational entropy expected from the volume mechanism is not observed.

One the of main advantages of first-principles calculations is the ability to perform thought experiments that help to understand a phenomenon. This will help us to settle the question of whether there is a volume effect in the Ni-Al system. The required thought experiment is deceptively simple. First disorder the material *at constant volume,* and then let it expand to its equilibrium volume. The changes in vibrational entropy associated with this two-step transformation are illustrated in Figure 3-9. In the second step, the expected vibrational entropy change expected from the volume mechanism is indeed observed. However, part of this entropy gain is cancelled by the *decrease* in vibrational entropy resulting from disordering the material at constant volume. This decrease is puzzling because it is in contradiction with the well-known "bond proportion" picture. The Ni-Al is a system with ordering tendencies: Disordering at constant volume should reduce the number of stiff Ni-Al bonds in favor of softer Ni-Ni and Al-Al bonds.

Just as in the Pd-V system, the effect of local relaxations resolves this apparent

Average bond lengths (Å)

Bond	$L1_2$	SQS-8	"Ideal"
Ni-Ni	2.468	2.550	2.425 (fcc Ni)
Ni-Al	2.468	2.444	2.463 (B2 NiAl)
Al-Al	–	2.680	2.813 (fcc Al)

Average bond Stiffnesses (eV/Å^2)

Bond	$L1_2$	SQS-8	"Ideal"
Ni-Ni	2.827	1.718	3.017 (from fcc Ni)
Ni-Al	2.567	2.884	2.219 (from B2 NiAl)
Al-Al	–	2.688	1.292 (from fcc Al)

Table 3.11: Average Bond Length and Bond Stiffness in the Ordered and Disordered States. The "ideal" bond length and stiffness are taken from structures which possess only one type of bond. Stiffness is defined as the spring constant associated with stretching a bond.

paradox. Ni and Al atoms have a very different size and significant relaxations may have to occur in order to allow them to coexist in a lattice. A very instructive way to look at this problem is to analyze the length and the stiffnesses of each type of bond in a structure where only one type of bond exists, so that nothing prevents this bond from reaching its "ideal" length. The Al-Al and and Ni-Ni are obviously found in the crystals of the corresponding pure elements, while the Al-Ni can be found in NiAl, which adopts the CsCl structure where all nearest neighbor bonds are identical. As seen in Table 3.11 the Ni-Ni and Ni-Al bonds have essentially the same "ideal" length. In the ordered $L1_2$ phase, which only possess Ni-Al and Al-Al bonds, each bond is essentially free to reach its "ideal" length. In the disordered state, the extremely long Al-Al bonds undergo a significant reduction in length (relative to pure Al) and their stiffness also correspondingly much higher than in pure Al.

In the thought experiment where the alloy is disordered at constant volume, the Al-Al would have to adopt a length that is close to 2.468 Å, just like the Ni-Ni and Ni-Al bonds. At that level of compression, the Al-Al would have stiffness[4] of 4.9 eV/Å^2. Thus, at this volume, the Ni-Ni and Al-Al bonds would be stiffer than the Ni-Al bonds. Disordering the alloy at constant volume therefore replaces the relatively soft Ni-Al bonds by stiffer Ni-Ni and Al-Al bonds, and the vibrational entropy decreases. The bond proportion model would incorrectly yield a increase in the vibrational entropy because it incorrectly assumes that Al-Al bonds have a unique stiffness, regardless of their lengths.

In order to show that bond length changes are really the key factor determining vibrational entropy changes in alloy, we perform the same analysis as in the Pd-V

[4]From the data of Table 3.11, we have:

$$\text{(Stiffness of Al-Al bond)} \approx 1.292 + (2.813 - 2.468)\frac{(2.688 - 1.292)}{(2.813 - 2.680)} = 4.913$$

system. We calculate the bond lengths and stiffnesses in a variety of structures and obtain a stiffness vs. length relationship (see Figure 3-10) through a least square fit.[5] This relationship is then used determine the bond stiffnesses in the $L1_2$ and the SQS-8 structures. The vibrational entropy difference obtained through this "bond stiffness vs. bond length" model is $0.03k_B$, which is in good agreement with the results of the more accurate calculations ($\pm 0.05k_B$).

3.5 Conclusion

The "bond proportion" and the volume mechanisms are clearly incapable of explaining the changes in vibrational entropy taking place in the Pd-V and Ni-Al systems when alloys undergo an order-disorder transition. The effect of local relaxations is so significant that it dominates the contributions of either two effects. Fortunately, the deviations from these two simplistic pictures can be very straightforwardly described by the "bond length vs. bond stiffness" picture.

Our results thus suggest a way to construct "transferable" force constants that would enable the calculation of the vibrational entropy of a large number of structures without having to recalculate force constants from *ab- initio* calculations for each of them. While the stiffness of a bond is unlikely to be transferable, the relationship between stiffness and length for a given type of chemical bond *is* transferable. Such transferable relationship can easily be determined by a fit to the force constants calculated from first-principles in a small set of structures. The vibrational entropy of any other structure could then be determined solely from the knowledge of its equilibrium geometry, an information that is already a byproduct of any *ab-initio* phase diagram calculation. This approach captures the essential physics determining vibrational entropy differences in alloys and presents an extremely promising way to include vibrational effects in phase diagram calculations at a moderate computational cost.

[5] Because the bond length changes are very large in the Ni-Al system, we allow for both a linear and quadratic dependence on bond length.

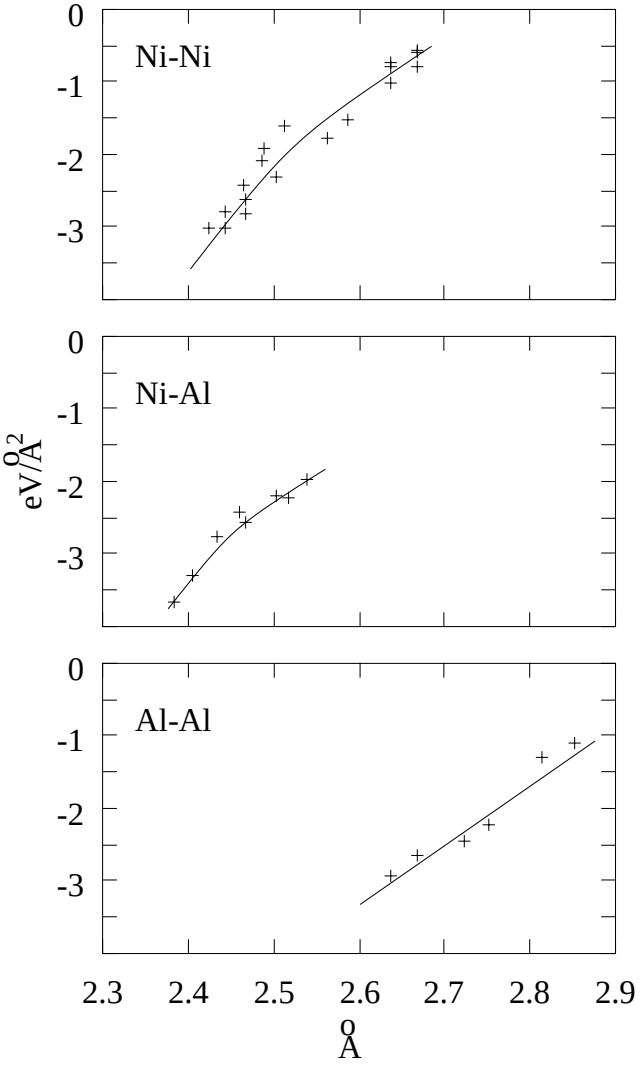


Figure 3-10: Bond Stiffness as a Function of Bond Length in the Ni-Al System. Only stretching terms are shown. Solid lines are the result of a fit to a second order polynomial. Each point corresponds to one type of bond in one of a set of fcc structures ($L1_2$, $D0_{22}$, SQS-8, fcc Al and fcc Ni, each taken at two different volumes)

Chapter 4

Correcting Overbinding in LDA Calculations

4.1 Introduction

Over the last three decades, first-principles calculations based on the local density approximation (LDA) of density functional theory (DFT) [64, 71] have been extensively used to successfully predict numerous materials properties [67]. While the LDA is known to have many limitations, we will focus here on properties that can, in principle, be exactly determined by DFT. These include any properties that can be derived from the knowledge of the electronic ground state energy for any given position of the nuclei.

One significant limitation of LDA is its "overbinding" of extended solids: Lattice parameters are typically underpredicted, while cohesive energies, phonon frequencies and elastic moduli are typically overpredicted. In this chapter, our main concern will be the biases in the equilibrium lattice constants, the phonon frequencies and the elastic moduli.

The Generalized Gradient Approximation (GGA) (see Ref. [108] and references therein), was introduced to address some of the weaknesses of the LDA. The GGA has been clearly shown to improve agreement with experiments for properties of finite systems such as atoms or molecules [108, 103]. However, the success of the GGA in extended systems has been more controversial:[77, 68, 45] despite many successes, [108, 105] GGA frequently overcorrects LDA's overbinding, sometimes yielding worse agreement with experiment than LDA. Obtaining accurate cohesive properties of crystalline phases from first-principles calculations thus remains a problematic issue.

This chapter describes how LDA's overbinding can be easily corrected for a large class of materials, by using a small amount of experimental input to correct the results of first-principles calculations. Our approach focuses on obtaining the correct equilibrium volume and relies on the fact that this correction is sufficient to significantly improve the accuracy of calculated elastic properties as well.

We start this chapter by giving a brief overview of Density Functional Theory. The remaining of the chapter is organized as follows. We first motivate the use of

the experimental equilibrium volume in LDA calculations and describe how LDA calculations can be combined with experimental input to obtain accurate equilibrium volumes. We then show, through the results of first-principle calculations, that errors in the equilibrium volume follow a trend that can be easily corrected with the use of a minimal amount of experimental input. We finally provide a formal explanation of the success of this simple approach and discuss the implication of our findings for the continuing search for better exchange-correlation functionals.

4.2 Density Functional Theory

Attempting to exactly solve the multibody Schrödinger equation of all the electrons in a system would be hopeless. The number of quantum states to consider grows exponentially with the number of particles and the resulting differential equations become quickly intractable. Fortunately, the focus of alloy theory (and solid state physics in general) is very often limited to determining the electronic ground state of a system. At most temperatures of interest, the width of the Fermi-Dirac distribution is small relative to the features of the electronic density of states in solid state systems and the occupation of the electronic states can be approximated reasonably well by a step function, which amounts to considering the system to be in its ground state.

The problem of finding the electronic ground state is a tremendously simpler task than to enumerate all eigenstates of an electronic system. As shown by Hohemberg and Kohn, the total energy of a system is a functional[1] of the electron density. Minimizing this functional with respect to the charge density yields both the true ground-state energy and its associated charge density. The advantage of such a formalism is that the charge density is a three dimensional object while the wavefunction of the ground state is a $3N$-dimensional object (in a N-electron system).

Kohn and Sham later showed that, for the purpose of finding the ground state energy, the multi-electron system represented by the Schrödinger equation can be replaced by a set of noninteracting single-electron equations, where the sum of the single-electron ground-state energies equal the ground-state total energy. In the framework of the so-called density functional theory (DFT), the coupling between the electrons are modeled by the adding an *exchange-correlation potential* term V_{xc} to the Hamiltonian of the non-interacting system:

$$\left(-\frac{\tilde{}^2}{2m_e}\nabla^2 + V(\mathbf{r}) + e^2\int\frac{\rho(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}d\mathbf{r}' + V_{xc}[\rho]\right)\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}) \qquad (4.1)$$

where $\mathbf{r}$ and $\mathbf{r}'$ are points in space, $V(\mathbf{r})$ is the external potential applied to the electron system, $\rho(\mathbf{r})$ is the charge density, $\psi_i(\mathbf{r}$ is the eigenfunction of state i and ϵ_i is its associated eigenvalue.[2] is and e, m_e and $\tilde{}$ are, respectively the electron charge,

[1] A *functional* is mapping from a function (here, the charge density) to a number (here, the energy).

[2] In DFT, the eigenvalues of the Hamiltonian have no direct physical meaning: They cannot be associated to the energy of eigenstate of the system.

the electron mass and Planck's constant.

Since this expression contains the charge density $\rho(\mathbf{r})$, which is determined by the wavefunctions $\psi(\mathbf{r})$, Equation (4.1) has to be solved self-consistently. A trial charge density is first used to find the wavefunctions using Equation (4.1). The resulting wavefunctions are used the calculate a charge density:

$$\rho\left(\mathbf{r}\right) = \sum_{i=1}^{N} \left|\psi_i\left(\mathbf{r}\right)\right| \tag{4.2}$$

where the sum is taken over the wavefunctions associated with the N lowest energy states (where N is the number of electrons). This charge density is then fed into Equation (4.1) and cycle is repeated until the desired level of accuracy is reached.

A nice feature of DFT is that, in principle, the exact exchange-correlation potential can be expressed as a functional of the charge density only (and not of the wavefunctions). The exact expression of $V_{xc}(\mathbf{r})$ is not known and determining it (or, at least, an accurate approximation of it) is one of the important challenges in solid-state physics. The local density approximation (LDA) to the exchange-correlation potential is by far the most commonly used. It consists in replacing the true exchange-correlation potential at point r by the exchange correlation potential of homogeneous electron gas of density $\rho(r)$. This simple approximation introduces various biases in the calculated properties of materials. This chapter describes how one of LDA's most obvious bias, its systematic underestimation of the lattice parameter, can be alleviated.

4.3 Method

4.3.1 The Importance of Volume

It has been previously observed [125, 33, 148, 147, 35] that, in many compounds, phonon frequencies calculated through first-principles are in closer agreement with experiment when the lattice parameters are artificially constrained to the experimental values. To determine whether this is a general trend, we performed a systematic investigation of the calculated bulk modulus in various elements, evaluated at the experimental lattice parameter. Since biases in the bulk modulus are usually correlated with biases in the phonon frequencies, investigating the bulk modulus provides a computationally simple way to investigate the bias in elastic properties in a large number of materials.

Figure 4-1 shows the result of our calculations, which will be described in detail in a later section. Using the experimental volume appears to essentially correct the bias in the calculated bulk modulus in a large class of materials. For all transition and noble metals tested, a substantial improvement is observed. More importantly, the magnitude of the correction is correlated with the magnitude of the error — indicating that this is clearly not just a random correction in the right direction. Interestingly, performing the same investigation with GGA shows that using the experimental volume for GGA calculations gives worse estimates of the bulk modulus than LDA at the

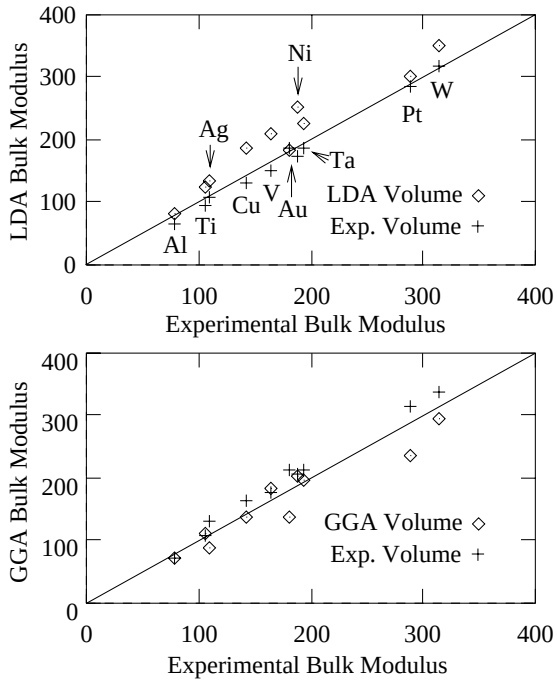


Figure 4-1: Calculated versus Experimental Bulk Modulus of Selected Elements. Evaluating the bulk modulus at the experimental volume considerably improves agreement with experiment. (The data for W and Ta was provided by the authors of Ref. [68].)

experimental volume (and also worse estimates than GGA at the GGA equilibrium volume). For this reason, we will focus on correcting the LDA equilibrium volume (instead of the GGA equilibrium volume).

Unfortunately, there are exceptions to this trend: for some elements, such as alkali metals, aluminum and silicon, evaluating the bulk modulus through LDA at the experimental lattice constant provides no improvement. The elements that appear to cause problems are those that have a small bulk modulus.

Nevertheless, we can conclude that evaluating elastic properties with LDA at the experimental volume provides significantly better agreement with experiment in a large class of materials including transition metals as well as other compounds studied by previous investigators. [125, 33, 148, 147, 35] A method of obtaining the correct lattice parameters from LDA calculations with as little experimental input as possible could therefore be extremely useful. This is the problem which we now turn to.

4.3.2 Conceptual Framework

In this section, we will first introduce a simple one-parameter correction to the LDA Hamiltonian of the system. This parameter is element dependent and can be tuned to obtain the true equilibrium volume. Such an approach avoids the conceptual problem associated with artificially fixing the volume without modeling the cause of the volume

change. We then describe how the knowledge of this parameter for simple systems enables us to extrapolate this parameter for systems where it is unknown. We finally present how we tested the applicability of our method in a large number of systems.

Let us define a structure to be a specific stable or metastable periodic arrangement of atoms of given types. Such a structure has well defined energy, equilibrium volume, elastic moduli and phonon frequencies. The fact that, in general, LDA correctly models elastic properties at the experimental volume suggests the following approximation for the total energy of a structure in the neighborhood of its equilibrium geometry:

$$E = E^{\mathrm{LDA}}(\{\mathbf{c}_i\}, \mathbf{a}_j) + \Delta E_{\mathrm{xc}}(\Omega)$$

where Ω is the average atomic volume, E^{LDA} is the energy obtained using LDA, for given atomic positions $\mathbf{c}_i$ and lattice vectors $\mathbf{a}_j$, while ΔE_{xc} is the correction introduced to model the effect of the non-local contribution to the exchange-correlation (which is not accounted for within LDA). The approximation introduced is to allow ΔE_{xc} to depend on the average atomic volume only. This approximation rests on the assumptions that LDA usually predicts reasonably accurate unit cell *shapes* (except for an isotropic scaling factor) and phonon frequencies, when evaluated at the experimental volume. If ΔE_{xc} were strongly dependent on the cell shape, the LDA cell shape would be significantly biased and if ΔE_{xc} were strongly dependent on atomic positions, the LDA would give incorrect estimates of the phonon frequencies.

We do not know ΔE_{xc} and seek to approximate it with a truncated Taylor series. Since we are interested in the equation of state in the neighborhood of the true equilibrium volume, we expand ΔE_{xc} in the neighborhood of the experimental volume Ω_{exp}:

$$\begin{aligned} E &= E^{\mathrm{LDA}}(\{\mathbf{c}_i\}, \mathbf{a}_j) + \Delta E_{\mathrm{xc}}(\Omega_{\mathrm{exp}}) + \\ &\quad + \frac{\partial \Delta E_{\mathrm{xc}}(\Omega_{\mathrm{exp}})}{\partial \Omega}(\Omega - \Omega_{\mathrm{exp}}). \end{aligned} \tag{4.3}$$

The truncation of the Taylor series after the first order can be justified from the fact that LDA seems to give accurate values of the bulk modulus when it is evaluated at the experimental volume, as mentioned before. In structures having a small bulk modulus, this assumption is more questionable, as the second order term is more likely to be non-negligible compared to the second derivative of E^{LDA}. This is probably the source of the inaccuracies in the bulk modulus we observed in soft elements.

While the term $\Delta E_{\mathrm{xc}}(\Omega_{\mathrm{exp}})$ may have an influence on cohesive energies, it leaves the equilibrium volume as well as elastic properties unaffected. We will thus focus on the quantity $\frac{\partial \Delta E_{\mathrm{xc}}(\Omega_{\mathrm{exp}})}{\partial \Omega}$ which has a direct impact on the equilibrium volume. It can be interpreted as the pressure that needs to be applied to the LDA solid so that its volume equals the experimental volume. This pressure, hereafter called the non-local exchange-correlation pressure (P_{xc}), is expected to be structure-dependent and we now turn to the problem of determining this pressure for a given structure using the minimum amount of experimental input.

The value of P_{xc} for each element, in its crystalline form, significantly changes from one element to the next. Elements that have large charge density inhomogeneities, like $3d$ metals, require a large corrective pressure while free-electron-like metals, such as alkali metals or aluminum, require a small pressure. It is often stated that LDA performs poorly in the case of alkali metals, since it significantly underpredicts their equilibrium volume. But the magnitude of the error introduced by the LDA, in terms of energy per unit volume (as measured by P_{xc}), is actually small. This small error in the energy translates into a large volume change only because alkali metals have such a low bulk modulus.

Obviously, using one P_{xc} per element, one can obtain the correct equilibrium volume for all elements. It is now interesting to check whether the value of P_{xc} in elements can help predict the value of P_{xc} in compounds. Remarkably, we have observed that taking the concentration-weighted average of P_{xc} of each element in the compound provides a simple and effective method to predict P_{xc} in a compound.

Since we are linearly interpolating the non-local exchange-correlation pressure, this naturally raises the following question: Why not linearly interpolate either the volume or the lattice parameter instead? The answer is that compounds that have the same composition can have slightly different volumes. For example, alloys typically expand by a fraction of a percent upon disordering. These volume relaxations are energetically significant and have an important impact on phonon frequencies [2, 136, 8] and mixing energies. Usually, LDA correctly predicts these relative volume changes and preventing them (by fixing the volume) would neglect an effect that LDA is actually able to model. Moreover, applying a pressure instead of fixing the volume has a firmer conceptual basis: Applying a pressure corresponds to introducing a perturbation to the Hamiltonian which causes a volume change. Fixing the volume imposes a volume change without modeling its cause.

4.3.3 Calculations

In order to benchmark this approach of linearly interpolating P_{xc}, we compute P_{xc} for a large number of compounds and compare it to the pressure interpolated from the P_{xc} of elements. The value P_{xc} can be determined by artificially constraining the volume to the experimental value and by calculating the pressure acting on the unit cell. For crystals of non-cubic symmetry, the cell shape is allowed to relax (at constant volume) until the pressure is isotropic. In all cases, we allow the internal degrees of freedom (i.e. the atom positions) to relax. This relaxation step is introduced so that the pressure is calculated in the same conditions as those in which it would be used, that is, without prior knowledge of cell shape and atomic positions.

The experimental volume used must be the one at absolute zero temperature. This is achieved by taking the experimental volume at room temperature (or higher) listed in Refs. [107, 102] and extrapolating it down to absolute zero using the thermal expansion data found in Ref. [130]. For intermetallics for which the thermal expansion down to absolute zero has not been determined experimentally, we use a concentration weighted average of the values of the pure elements. For compounds that exhibit an allotropic or a magnetic transformation between the temperature at which the

volume is measured and absolute zero, we approximate the thermal expansion of the high-temperature phase at low temperature by the thermal expansion of the low-temperature phase and disregard the volume change taking place at the phase transition. Note that, in most compounds, these corrections are quite small (typically 0.2% change in the lattice constant) compared to the LDA error (1% to 2%). Hence, any inaccuracies in these extrapolations have a small impact on our results. In fact, calculations at the room temperature volume exhibit essentially the same behavior.

Our calculations are performed with the VASP [75, 74] package which uses ultra-soft [139] pseudopotentials [113] and a plane-wave basis. The equilibrium volume can be sensitive to the quality of the pseudopotentials used. We thus verified that the equilibrium lattice constants we obtain do not differ significantly [3] from previous full-electron calculations in elements [68, 105, 108, 45]. The energy cut-off used ranges from 200 eV to 360 eV depending on the elements present in the structure. For ultra-soft pseudopotentials, these cut-offs are largely sufficient to obtain pressures with a precision of about 1% for transition metals and simple metals. To minimize the effect of Pulay stress, calculations are performed at a sequence of various volumes ranging from -18% to $+12\%$ around the experimental volume, at a constant energy cut-off. At each volume, all degrees of freedom are fully relaxed and the resulting energies are fitted to a polynomial of degree 4. The resulting equation of state is then used to obtain the quantities of interest. The Brillouin zone is sampled using Monkhorst-Pack special points and integrated with the help of the tetrahedron method with Bloech corrections. [4] For all structures, the total number of k-points in the Brillouin zone is chosen to be around $(15)^3$ divided by the number of atoms in the unit cell. LDA calculations relied on the Ceperley-Alder exchange-correlation [29], as parametrized by Perdew and Zunger [109], while GGA calculations relied on the Perdew-Wang '91 functional [108]. In all calculations, the same exchange-correlation functional is used for both the pseudopotential generation step and the actual calculations. Scalar-relativistic effects are accounted for in the pseudopotential. Non-spin polarized calculations are used for all compounds which are paramagnetic at room temperature. Spin-polarized calculations are used for metallic nickel, the only magnetic compound considered here.

4.4 Results

We choose binary systems where the number of well characterized phases is sufficiently large to provide a stringent test of the linear dependence of P_{xc} on concentration. Systems exhibiting intermetallic phases are ideal tests for our method, as the known structures are well characterized experimentally and have small unit cells, which makes calculations easily tractable.

[3] Lattice constants of Ag, Al (LDA), Au, Cu, Na, Ni, Pd, Pt, Si, and V differed by less that 0.6%. Lattice constant of Ti (LDA and GGA) and Al (GGA) differed by about 1%.

[4] For relaxation steps the Methfessel-Paxton smearing method of order 1 with a width of 0.1 eV is used instead, to avoid the problem that accurate forces cannot be obtained with the tetrahedron method.

The non-local exchange-correlation pressures for all the intermetallic phases tested are shown in Figure 4-2. In this figure, the accuracy of the volume determined with the help of an exchange-correlation pressure relative to the LDA volume can be noted by observing that the data points are much closer to the interpolating line than to the zero axis. It is remarkable that the linear relationship holds even when the phases do not share a common parent lattice. [5] In particular, the different crystalline structures of titanium exhibit almost the same P_{xc}. This is also the case with lithium, whose hcp and bcc phases have a non-local exchange-correlation pressure of -1.48 GPa and -1.39 GPa, respectively.

The same linear relationship appears to hold for ionic and covalent solids as well, as shown in Figures 4-3 and 4-4. While the agreement is less satisfactory in the case of alkali halides, our method still reduces LDA's error on the equilibrium volume by 50%. The case of the ionic compounds is a particularly stringent test of our method because the pure elements used to predict the pressures for the compounds are chemically very different from these compounds: on one side, they are metals (Ca, Mg, Na, K, and Cs), while on the other side, they are covalently bounds molecules held together by van der Waals forces (O_2 and Cl_2).

Our method may appear less successful in the case of the covalent crystal SiC. However, our method predicts the lattice constant of SiC within 0.46%, which is more accurate than both the LDA lattice constant (-0.85% error) and the lattice constant obtained by taking either the mean of the lattice parameter of C and Si (3.2% error) or the mean of their molar volume (7.4% error).

Unfortunately, our linear interpolation scheme does fail in some systems. During our benchmarking, we found three systems where this approach was unsuccessful: In the Ga-As and Li-Co-O systems, the pressure was clearly not linear in concentration, while in the case pure carbon, the pressure associated with diamond and graphite were radically different. In all cases, the failure is associated with the presence a markedly anisotropic bonding in some of the compounds, which makes the use of an isotropic pressure inappropriate. The value of the stress in graphite, when calculated with the LDA at the experimental lattice parameters, best illustrates the nature of the problem. While the stress perpendicular to the graphitic layers (0.09 GPa) differs from the isotropic component of the stress in the diamond structure (14.4 GPa), the stress along the graphitic layers (16.0 GPa) agrees very well with the value obtained in diamond.

We have up to now only focused on predicting the lattice parameters, but in principle, our correction to the LDA Hamiltonian can just as well be used to improve the accuracy of calculated formation energies. As an additional test, we now show that in two cases where LDA is known to predict the wrong ground state, our correction to the LDA Hamiltonian is sufficient to obtain the correct ground state.

If our correction to the LDA is indeed a function of volume and concentration

[5] The parent lattice is defined as the set of all atomic positions in the crystal, disregarding the type of atom occupying each site. As is commonly done, we also disregard small relaxation away from the ideal lattice positions when labeling parent lattices. For example, the $L1_0$ structure is considered to have an fcc parent lattice despite its tetragonal distortion.

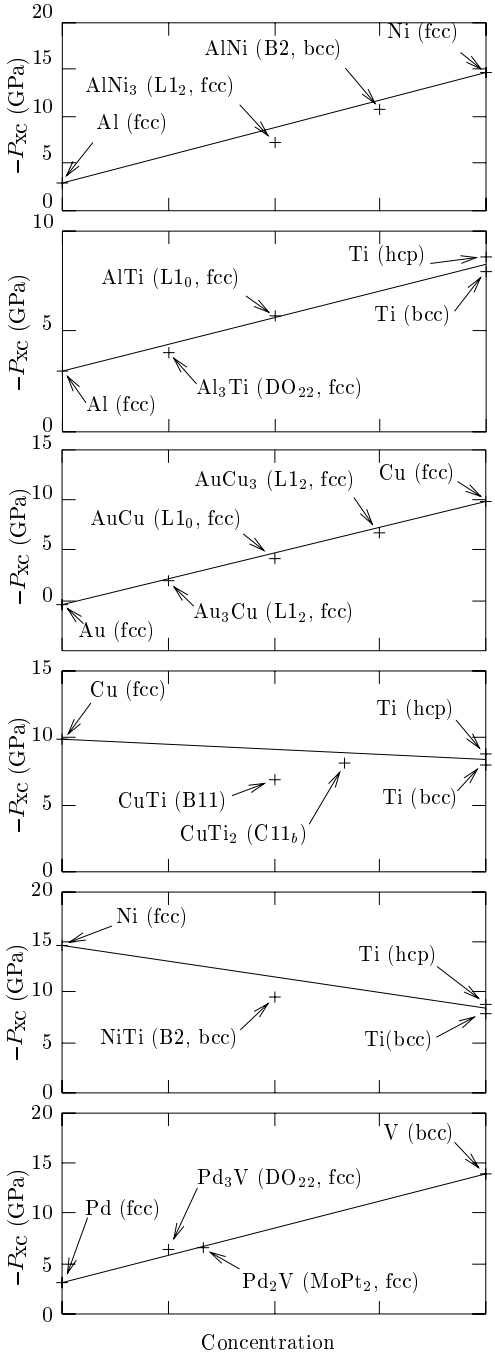


Figure 4-2: Non-Local Exchange-Correlation Pressure as a Function of Concentration in Metallic Systems.

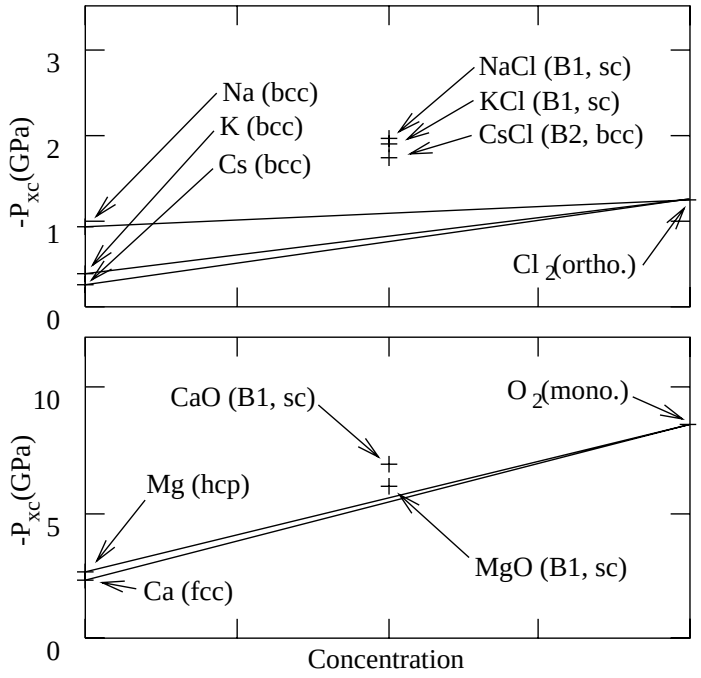


Figure 4-3: Non-Local Exchange-Correlation Pressure as a Function of Concentration in Ionic Systems. The pressures for oxygen and chlorine are obtained from their solid crystalline structure at low temperature (23K for oxygen and 100K for chlorine).

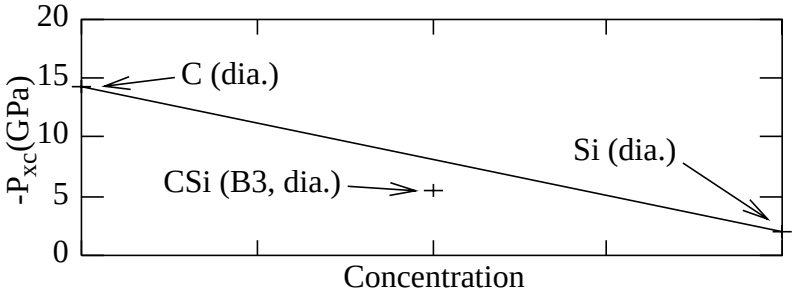


Figure 4-4: Non-Local Exchange-Correlation Pressure as a Function of Concentration in SiC, a Covalent System.

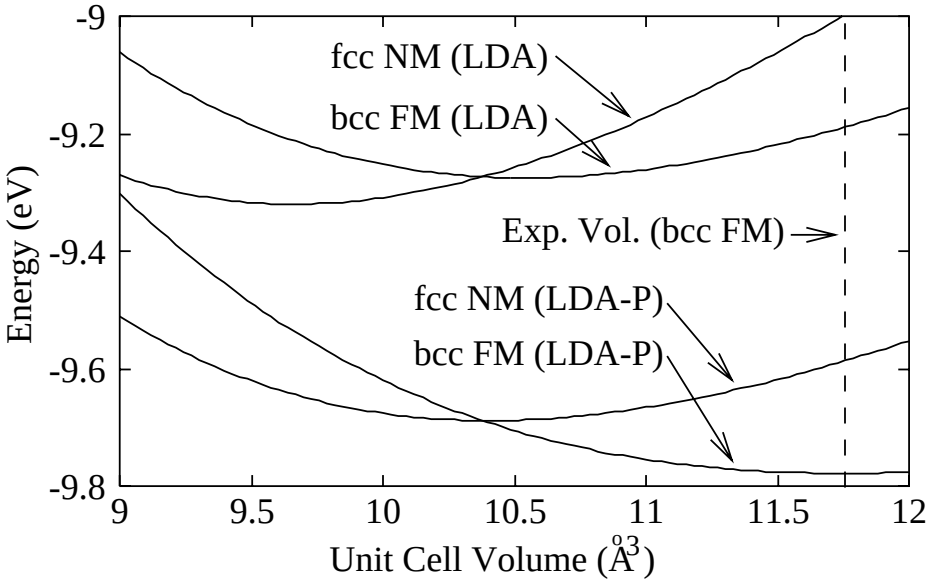


Figure 4-5: Equation of State of Iron Calculated with the LDA only and with the LDA Using an Exchange-Correlation Pressure (LDA-P).

only, the correction to the energy of two structures which share the same composition should be of the same form. It follows that the volume-independent terms in Equation (4.3) do not affect the relative stability of the two structures. The volume dependent part, $P_{\mathrm{xc}}\Omega$, which does affect the relative stability, can be determined as before, from the pressure calculated with the LDA at the known experimental volume of one of the structures. The same quantity $P_{\mathrm{xc}}\Omega$ is then added to the equation of state of both structures.

One of LDA's most notorious failures is its prediction that the ground state of iron is a non-magnetic fcc structure instead of the observed bcc ferromagnetic structure. As shown in Figure 4-5, the problem is readily corrected by adding a term of the form $P_{\mathrm{xc}}\Omega$ determined from the knowledge of the experimental lattice parameter of bcc iron.

Another example is the $LiMnO_2$ compound used in lithium ion batteries[83] whose ground state is known to be an orthorombic antiferromagnetic structure. LDA incorrectly predicts a so-called layered structure to be the ground state. Once again, our simple correction, based on the knowledge of the true cell volume of the orthorombic structure, restores the correct ground state (see Figure 4-6).

4.5 Discussion

Our observations give us more than a practical way to correct LDA's overbinding. They also provide useful information regarding the structure of the non-local contribution to the exchange-correlation energy. Our results indicate that a large part of the error on the total energy obtained through LDA takes the form of a linear function of the volume, where the constant of proportionality is itself linear in concentration.

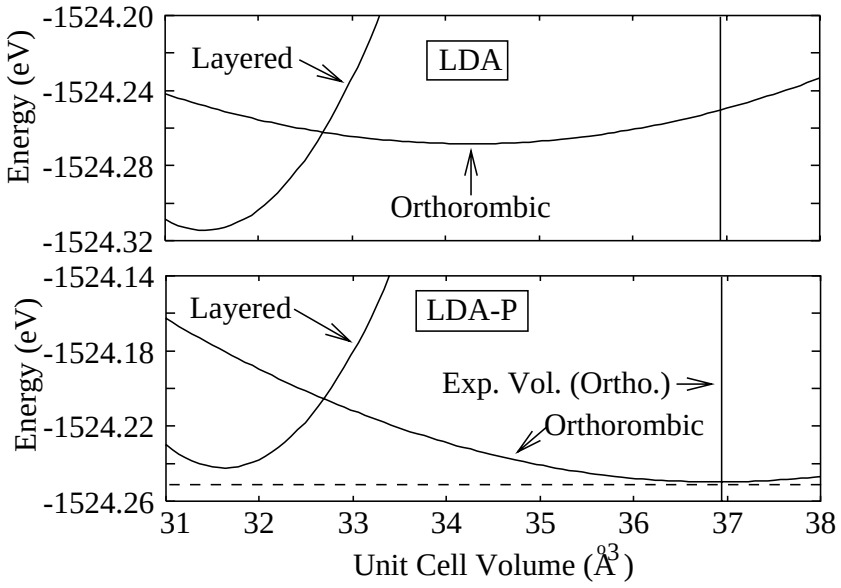


Figure 4-6: Equation of State of $LiMnO_2$ Calculated with the LDA Only and with the LDA Using an Exchange-Correlation Pressure (LDA-P).

Only very specific corrections to the LDA exchange-correlation energy can give rise to such a contribution. This section is thus devoted to determining the most general form of exchange-correlation functional that is compatible with our observations. We will motivate such a functional form based on the results of previous calculations as well as on reasonable physical assumptions.

Admittedly, our description of the non-local contribution to the exchange correlation energy is highly simplified: It is not intented to model strong electron correlation effects. We rather seek to remove LDA's overbinding problem, which is present in all materials, even when electron correlations effects are relatively weak. Nevertheless, we believe that this analysis provides important insight into the continuing search for more accurate exchange-correlation functionals.

This discussion is organized as follows. We first determine which form the correction to the LDA exchange-correlation functional must take in order to give rise to an energy contribution that is a linear function of volume only. We will then examine the implications of constraining the exchange-correlation pressure to be linear in concentration. In particular, we will argue that this behavior is possible only if the correction to the LDA mainly arises from the large non-uniformity in the charge density near the atom nuclei. This picture naturally provides a physical interpretation to the exchange-correlation pressure in terms of the zero-order coefficient of the functional expansion of the exchange-correlation energy, when the point of expansion is taken to be the atomic charge density.

4.5.1 Linearity of $\Delta E_{\mathbf{xc}}$ in Volume

We can divide the volume of the solid in two parts: "core" regions where the charge density is essentially identical to the charge density of a free atom, and "interstitial" regions where this approximation ceases to apply. The core regions are simply considered to be spheres centered on each nuclei. The interstitial regions include all points outside of these spheres. While the charge density in the core regions is clearly non-uniform, it is essentially frozen. Hence, even if the LDA is a poor approximation in the core regions, the error it introduces is essentially constant and has little influence on cohesive properties and in particular, on the equilibrium volume. We therefore seek the source of LDA's volume underestimation in the inaccurate description of the exchange-correlation energy in the interstitial region. Our approach is to correct the exchange-correlation energy in the interstitial regions only, leaving the correction to be made in the core regions unspecified. We simply need to assume that the exchange-correlation potential correction in the core regions is independent of the environment of the core and is a relatively smooth continuation of the correction in the interstitial regions, in order not to dramatically affect the wavefunctions in the interstitial region.

In the remainder of this section, we will describe how our results lead to the conclusion that most of LDA's bias can be corrected by adding to the LDA exchange-correlation energy per unit volume $e_{\mathrm{xc}}^{LDA}(\mathbf{r})$ in the interstitial region, a term of the form

$$\Delta e_{\mathrm{xc}}(\mathbf{r}) = \Delta e_{\mathrm{xc}} + \Delta \mu_{\mathrm{xc}} \rho(\mathbf{r}), \tag{4.4}$$

where $\rho(\mathbf{r})$ is the charge density at point $\mathbf{r}$ and Δe_{xc} and $\Delta \mu_{\mathrm{xc}}$ are corrections to the exchange-correlation energy and potential. We take Δe_{xc} and $\Delta \mu_{\mathrm{xc}}$ to be uniform over the interstitial region for the following reasons.

The observation that our non-self-consistent correction successfully predicts cohesive properties indicates that the correction to the exchange-correlation potential is relatively uniform over the interstitial region, or at least that the deviations away from a uniform correction are sufficiently small to allow a first-order perturbative treatment (*i.e.* a non-self-consistent correction). Any large non-uniformities in the correction to the exchange-correlation potential would inevitably lead to large corrections to the LDA wavefunctions, leaving the success of our non-self-consistent approach unexplained.

Prior investigations also support the validity of this non-self-consistent approach. The error in the LDA exchange-correlation potential was shown to be surprisingly uniform in systems where the exact exchange-correlation potential can be obtained at a moderate computational cost: in light atoms [7] and in light atoms dimers [60]. In these systems, one can see that a simple shift is sufficient to dramatically improve the accuracy of the calculated exchange-correlation potential in the region that would correspond to the interstitial region in extended systems. A non-self-consistent approach can also be justified from the fact that GGA, when implemented in a non self-consistent way by simply using the GGA exchange-correlation functional

with the LDA charge density as an input, yields results that are very similar to a fully self-consistent procedure.[48, 73] The magnitude of GGA's corrections over LDA results are comparable to the magnitude of the corrections devised in the context of the present work and we thus expect the perturbative approach to be appropriate.

A uniform shift in the exchange-correlation potential is modeled by taking $\Delta\mu_{\text{xc}}$ to be uniform over the interstitial region and independent of the interstitial charge density, which prohibits the inclusion in Equation (4.4) of higher-order terms in ρ. Note that the value of $\Delta\mu_{\text{xc}}$ has little effect on the cohesive properties: to the extent that the charge density in the core region is truly frozen, the total charge in the interstitial region is also constant and the integral of $\Delta\mu_{\text{xc}}\rho(\mathbf{r})$ over the interstitial regions is volume-independent.

On the contrary, the shift in the exchange-correlation energy per unit volume Δe_{xc} directly gives rise to the exchange correlation pressure P_{xc} and, in fact, is numerically equal to it. The fact that the exchange-correlation pressure for structures which have the same composition but do not share a common parent lattice is nearly identical constrains the correction to the exchange-correlation energy to be uniform in the interstitial region. Since the shape of the interstitial space strongly depends on the type of lattice, any non-uniformity would make the P_{xc} lattice-dependent. A convenient consequence of the uniformity of Δe_{xc} is that the exchange-correlation pressure is independent of the precise radius chosen to delimit the core regions.

4.5.2 Importance of the Core Charge Density

Using the observation that most of LDA's error on the total energy takes the form of a correction that is linear in volume enabled us to considerably restrict the type of correction to the LDA exchange-correction functional we are focusing on. Using the fact that the exchange-correlation pressure is linear in atomic concentration lets us be more specific about what determines the unknown constants Δe_{xc} and $\Delta\mu_{\text{xc}}$. We have already indicated that these constants should be essentially independent of the interstitial charge density. Analyzing the linearity of P_{xc} with concentration will make it clear that these constants are essentially determined by the charge density in the *core* regions.

Limiting ourselves to the conservative assumption that $\Delta e_{\text{xc}}(\mathbf{r})$ at a point $\mathbf{r}$ in the interstitial region only depends on the charge density in the interstitial region (as done in the GGA, for example), makes it difficult to explain our simple linear relationship between P_{xc} and concentration. First, the charge density in the interstitial region is highly dependent on the exact location of neighboring atoms — information the concentration by itself seems unable to convey. Second, even if we assume that atoms sit exactly on the sites of a parent lattice common to all phases in the system of interest *and* approximate the charge density in the interstitial regions by overlapping atomic charge densities, we still face a problem. The exchange-correlation functional is a non-linear function of the charge density and summing atomic charge densities does not translate into summing the exchange-correlation contributions coming from each neighboring atom. The only possibility is that Δe_{xc} can be approximated by a linear functional. But then, we obtain a ΔE_{xc} that is volume-independent since, as

we have indicated before, the total interstitial charge density is essentially volume-independent.

However, if we consider the possibility that $e_{\rm xc}(\mathbf{r})$ at a point in the interstitial region depends on the charge density in a neighborhood large enough to include the core regions of the neighboring atoms, we will see that a simple explanation for our findings emerges. While the dependence of $\Delta e_{\rm xc}$ on the interstitial charge density needs to be assumed linear, no such restriction needs to be imposed in the case of the dependence on the core charge density. Assuming that our volume-dependent correction to the LDA exchange-correlation energy in the interstitial region does not arise from the small non-uniformities of the interstitial charge density, but rather from the large non-uniformities of the core charge densities, avoids most of the difficulties mentioned previously. First, the charge density in the core regions is independent of the exact location of neighboring atoms. Second, the problem of the non-linearity of the exchange-correlation energy is irrelevant since, the core charge density is not obtained by a sum of the contributions of neighboring atoms but by the contribution of one atom only. The only assumption needed is that the correction arising from each neighboring atom are additive, which we will motivate shortly.

Previous calculations in systems simple enough to allow an accurate evaluation of the exchange-correlation energy provide substantial evidence that the core charge density has an important effect on the exchange-correlation energy in the interstitial region.

The spatial extent of the exchange-correlation hole is a measure of the size of the neighborhood which is expected to influence the exchange-correlation energy at a point. Accurate calculations of the exchange-correlation hole in real systems show that it extends up to distances that are comparable to the interatomic distances. This has been directly observed in calculations in silicon [65], while calculations of the electron pair-correlation function in diamond [44] and metallic lithium [127] indicate that a similar feature is present in those systems.[6] Additionally, the radius of the exchange-correlation hole obtained through LDA, $\left\langle \frac{1}{\mathbf{r}_{\rm xc}} \right\rangle = \epsilon_{\rm xc}^{LDA}$, provides an estimate of the spatial extent of the true exchange-correlation hole. For all the metals investigated in this work, this approximate exchange-correlation hole radius is comparable to the distance separating neighboring atoms.

Another indication of the influence of the core charge density on the interstitial exchange-correlation energy is that the exchange-correlation hole exhibits a surprising feature when a reference electron is located in the neighborhood of an isolated atom or a molecule. The exchange-correlation hole possesses not only a peak centered on the reference electron but also exhibits additional peaks where the exchange-correlation hole overlaps the core of the neighboring atoms, as illustrated in Figure 4-7. Whenever the reference electron is located in regions of relatively low electronic density, the exchange-correlation hole tends to "leak" to the nearest region where the charge density is large. This effect is especially large for isolated atoms [43], molecules [138] and surfaces [76]. This is responsible for the incorrect asymptotic behavior of

[6] The electron pair-correlation function is directly related to the exchange-correlation hole [67].

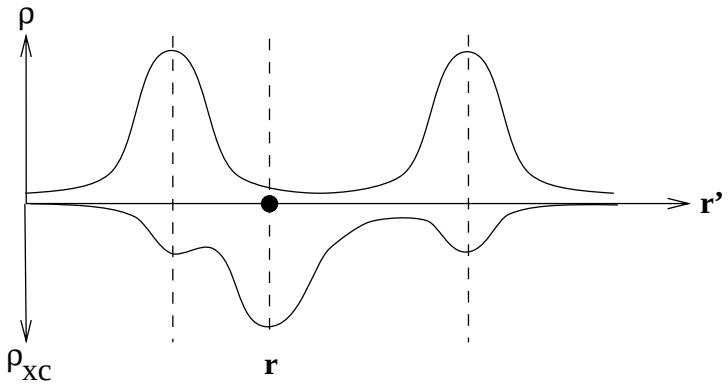


Figure 4-7: Schematic Shape of the Exchange-Correlation Hole ($\rho_{\rm xc}$) for a Reference Electron Located at Point $\mathbf{r}$ in the Interstitial Region. The exchange-correlation hole exhibits extra peeks where the hole overlaps the core charge density (peaks in the graph of ρ) of neighboring atoms.

the LDA exchange-correlation as the reference electron is moved to infinity [138]. LDA predicts an exchange-correlation hole which follows the reference electron as it is moved away, whereas the true exchange-correlation hole actually largely remains in the atom, the molecule, or near the surface. This is clearly an indication that regions of large charge density can in principle have an impact on the exchange-correlation energy up to arbitrarily large distance. Accurate Quantum Monte Carlo calculations have shown a similar effect in extended solids. Calculations in silicon [65] show the presence of these multiple peaks while calculations on metallic lithium [127] and in diamond [44] show multiple peaks in the electron pair correlation function from which multiple peaks in the exchange-correlation hole can be inferred.

4.5.3 Linearity of $P_{\rm xc}$ in Concentration

While we have motivated that $\Delta e_{\rm xc}$ in the interstitial region is mainly determined by the charge density in the cores we still have to describe how a linear dependence on concentration arises.

Let us first define a few quantities. The exchange-correlation energy per unit volume $e_{\rm xc}(\mathbf{r})$ at a point $\mathbf{r}$ can be expressed in terms of the electron pair-correlation function:

$$e_{\rm xc}(\mathbf{r}) = \frac{1}{2}\int g(\mathbf{r},\mathbf{r}')d^3\mathbf{r}', \tag{4.5}$$

where $g(\mathbf{r},\mathbf{r}')$ is a functional of the charge density that embodies all the information regarding the electron pair-correlation function. More specifically,

$$g(\mathbf{r},\mathbf{r}') = \frac{1}{|\mathbf{r}-\mathbf{r}'|}\rho(\mathbf{r})\left(\overline{h}(\mathbf{r},\mathbf{r}')-1\right)\rho(\mathbf{r}')$$

where $\overline{h}(\mathbf{r},\mathbf{r}')$ is the coupling-constant-averaged electron pair-correlation of the sys-

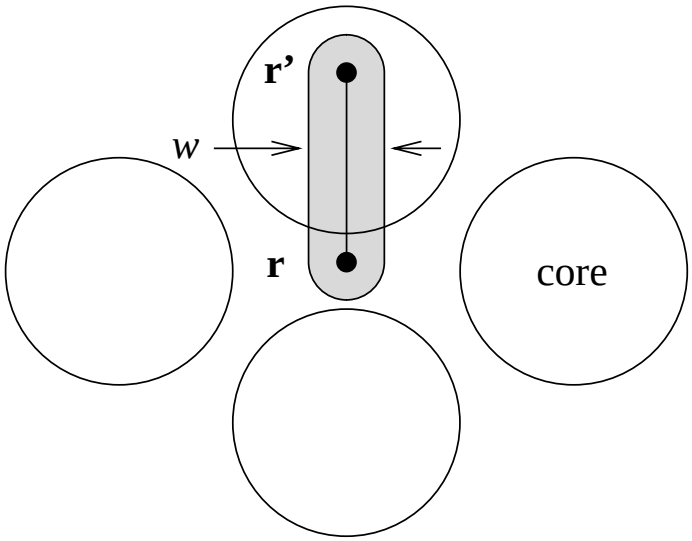


Figure 4-8: Idealized Shape of the Region Whose Charge Density is Expected to Influence the Electron Pair Correlation Function Between Points $\mathbf{r}$ and $\mathbf{r}'$.

tem, as defined, for example, in Reference [72]. (The explicit functional dependence of $e_{\rm xc}(\mathbf{r})$ and $g(\mathbf{r},\mathbf{r}')$ on the charge density ρ has been omitted to simplify the notation).

The difference between the true and the LDA exchange-correlation follows a similar structure:

$$\Delta e_{\rm xc}(\mathbf{r}) = \frac{1}{2}\int \Delta g(\mathbf{r},\mathbf{r}')d^3\mathbf{r}', \tag{4.6}$$

where $\Delta g(\mathbf{r},\mathbf{r}')$ is defined as the difference between the true value of $g(\mathbf{r},\mathbf{r}')$ and its approximation obtained with the LDA.

With these definitions in hand, we now introduce a plausible restriction to the functional dependence of $\Delta g(\mathbf{r},\mathbf{r}')$ on the charge density which guarantees that the shift in the exchange-correlation energy density in the interstitial region is linear in concentration. We consider $\Delta g(\mathbf{r},\mathbf{r}')$ to be a functional of the charge density in a relatively narrow region surrounding the segment joining points $\mathbf{r}$ and $\mathbf{r}'$. The width w of this region is assumed to be small relative to interatomic distances, as illustrated in Figure 4-8.

Under this assumption, the integration of $\Delta g(\mathbf{r},\mathbf{r}')$ over $\mathbf{r}'$ in Equation (4.5) is a linear operation: $\Delta e_{\rm xc}(\mathbf{r})$ reduces to a sum of integrals over disjoint sectors S_i (see figure 4-9), each of which representing the contribution coming from one particular neighboring atom i:

$$\Delta e_{\rm xc}(\mathbf{r}) = \sum_i \int_{S_i} g(\mathbf{r},\mathbf{r}')d^3\mathbf{r}'$$

If w is not too large compared to interatomic distances, there will be only a small coupling between the charge density of neighboring sectors and a term corresponding

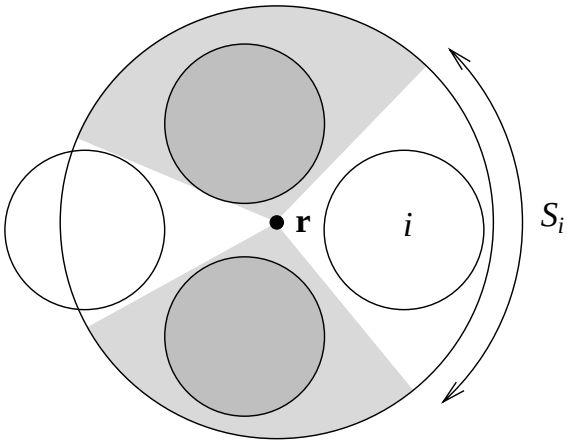


Figure 4-9: The exchange-correlation energy at a point $\mathbf{r}$ in the interstitial region can be written as a sum of integrals over disjoint sectors S_i, each of which representing the contribution coming from one particular neighboring atom i.

to sector S_i will mainly depend on the charge density in sector S_i.

These approximations allow the difference between the LDA exchange-correlation energy per electron and the true one, $\Delta\epsilon_{\rm xc}$, in the interstitial region to be approximated by a sum of contributions coming from nearby atoms. The linearity of $\Delta\epsilon_{\rm xc}(\mathbf{r})$ in the atomic concentration is thus a consequence of the well-known fact that only the spherical average of the exchange-correlation hole plays a role in determining the exchange-correlation energy.

4.5.4 Atom-Specific Correction

We can now give a simple interpretation to the constants $\Delta e_{\rm xc}$ and $\Delta\mu_{\rm xc}$ introduced earlier. Since the correction to the LDA exchange-correlation in the interstitial region can be approximated by a sum of *independent* contributions from neighboring atoms, we can approximate the correction to the exchange-correlation energy at a point in the interstitial region by replacing the exact contributions coming from each neighboring atom by the contributions of isolated atoms centered at the same locations, as illustrated in Figure 4-10. Even though this procedure may seem approximate, including non-local effects, even in this simplified way, is an improvement over not including them at all.

Consider an isolated neutral atom i centered at the origin with charge density $\rho_i(\mathbf{r})$. This atom has a non-zero non-local exchange-correlation energy density $\Delta e^i_{\rm xc}(\mathbf{r})$ at every point in space, which can be integrated over space to give its total non-local exchange-correlation energy:

$$\Delta E^i_{\rm xc} = \int \Delta e^i_{\rm xc}(\mathbf{r}) d^3\mathbf{r}.$$

The quantity $\Delta E^i_{\rm xc}$ will of course change if the charge density in the neighborhood of

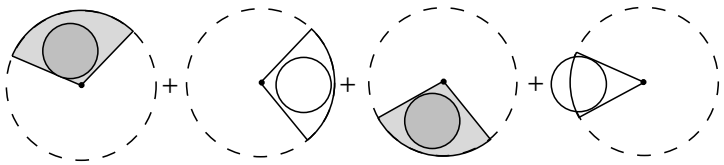


Figure 4-10: Approximating the Non-Local Exchange-Correlation Energy as a Sum of Independent Contributions

this atom is perturbed by an amount $\delta\rho(\mathbf{r})$. Provided that $\delta\rho(\mathbf{r})$ is not too large, we can approximate the true dependence of $\Delta E^i_{\rm xc}$ on $\delta\rho(\mathbf{r})$ by an expansion of the true exchange-correlation function in terms of a series of homogeneous functional of $\delta\rho(\mathbf{r})$ of degree 0, 1, 2 ...

The zero order term of this expansion is simply the value of $\Delta E^i_{\rm xc}$ at the "point" of expansion, that is, evaluated at the atomic charge density $\rho_i(\mathbf{r})$. Thus the zero-order correction to LDA consists in adding the energy $\Delta e^i_{\rm xc}(\mathbf{r})$ at every point in the neighborhood of atom i. In accordance with our earlier assumptions we consider the dependence on $\mathbf{r}$ to be weak outside of the core. The shift in exchange-correlation energy in the interstitial region introduced earlier is thus simply the sum of the atom-specific shift of the neighboring atoms:

$$\Delta e_{\rm xc} = \sum_i \Delta e^i_{\rm xc}$$

The first order term of this functional expansion can be expressed in terms of the functional derivative of the non-local exchange-correlation energy evaluated at the atomic charge density:

$$\begin{aligned} \delta\Delta E^i_{\rm xc} &= \int \left.\frac{\delta\Delta E^i_{\rm xc}}{\delta\rho}\right|_{\rho=\rho_i} \delta\rho(\mathbf{r})d\mathbf{r} \\ &\equiv \int \Delta\mu^i_{\rm xc}(\mathbf{r})\delta\rho(\mathbf{r})d\mathbf{r}, \end{aligned}$$

which defines an atom-specific correction to the exchange-correlation potential $\Delta\mu^i_{\rm xc}(\mathbf{r})$. As before, assuming a weak dependence on $\mathbf{r}$ and summing the contributions of neighboring atoms provides the shift in the exchange-correlation potential:

$$\Delta\mu_{\rm xc} = \sum_i \Delta\mu^i_{\rm xc}$$

In the same spirit as in the beginning of this section, higher order terms of the functional expansion are neglected.

The non-local exchange correlation pressure ($P_{\rm xc}$) is thus more than a fitting parameter in an empirical equation of state. Under plausible assumptions, this pressure can be interpreted as the sum of the zero-order corrections to the LDA exchange-correlation energy of each atom present in a given compound.

4.6 Conclusion

This chapter's first objective is to emphasize that LDA's overestimation of phonon frequencies and elastic moduli is essentially a consequence of its underestimation of the equilibrium volume. In a large class of materials, calculated elastic properties are in good agreement with experiment when the LDA calculations are performed at the experimental volume.

Given the importance of obtaining the correct equilibrium volume, we propose a simple and yet effective way to correct LDA's bias: apply a negative pressure such that the equilibrium volume agrees with experiment. We have argued that this so-called non-local exchange-correlation pressure originates from the non-local contribution to the exchange-correlation energy per unit volume in the interstitial region of the solid.

This method would be of limited usefulness if one needed a different non-local exchange-correlation pressure for each compound. Fortunately, we have observed that the non-local exchange-correlation pressure of a compound can be accurately determined by taking the concentration-weighted average of the non-local exchange-correlation pressure of elements. This linear relationship holds for all the metallic, ionic and covalent systems we have investigated and we are confident of its wide applicability.

We then propose a simple explanation for this surprising linear behavior which relies on the following main assumptions.

- The non-local contribution to the exchange-correlation energy in the interstitial region mainly originates from the large non-uniformity of the core charge density and not from the weak non-uniformities of the interstitial charge density.
- This non-local correction can be approximately expressed as a sum of independent contributions coming from nearby atoms.
- These contributions take the form of a nearly uniform, atom-specific, and charge-density-independent correction to the exchange-correlation energy density in the interstitial region.

These assumptions, which are crucial in obtaining a linear dependence between the non-local exchange-correlation pressure and concentration, are guided by the results of previous investigations which have determined the exact exchange-correlation energy in simple systems.

Our results have implications at two different levels:

- A simple linear interpolation scheme provides a simple and accurate way to correct LDA's volume underestimation. Furthermore, this correction is often sufficient to dramatically improve the accuracy of calculated elastic properties.
- The fact that such a simple scheme performs so well provides strict constraints on the behavior of the non-local corrections to the LDA and provides helpful clues in the continuing search for better exchange-correlation functionals. Atom-specific corrections determined from isolated-atom calculations would appear a promising approach.

Chapter 5

The ground state problem

5.1 Introduction

The topology of a phase diagram is mainly determined by the ground states of the alloy system considered, that is, the structures that have the lowest energy (at a given chemical composition). In a system where the only ground states are the pure elements, the phase diagram will consist of a simple miscibility gap.[1] In a system exhibiting various intermetallic phases, each ground state manifests itself as a line compound (with sometimes a relatively wide concentration range) that eventually disorders or melts as temperature increases. The ground states are the structures that are the most stable at the absolute zero of temperature and this stability usually persists at higher temperatures as well.

In this chapter we will first introduce the basic framework that allows the determination of the ground states of an alloy. We focus on the case where all ground states are assumed to be superstructures of a known parent lattice (*e.g.* fcc or bcc). (If needed, more than one parent lattice could be considered in turn, and the resulting ground states can be combined.) We will then describe the key concepts that are required to improve existing ground state search techniques, before introducing a new approach to solve the ground state problem. We finally go through a step-by-step example of the use of this method in a simple case in order to illustrate its usefulness.

5.2 Formalism

5.2.1 Formulation of the ground state problem as a minimization problem

The so-called ground state problem is the apparently impossible task of looping through the infinite set of every possible structure (which can have arbitrarily large

[1] The critical temperature may be sufficiently low so that perfect solid solubility is observed down to very low temperatures. Nevertheless, the third law of thermodynamics requires a system that is in equilibrium to eventually phase separate as its temperature approaches absolute zero.

unit cells) in search for the structure that minimizes the energy at a given composition. Fortunately, the basic formalism of the cluster expansion permits a very elegant way to determine the ground states of an alloy system. In the cluster expansion formalism, the energy of the alloy is expressible as a linear function of the correlations ρ_α (see Section 2.2.1):

$$\frac{E}{N} = \sum_\alpha m_\alpha J_\alpha \rho_\alpha.$$

Thus, in order to find the minimum energy structures, one can replace the search through the set of structures by a search through the set of correlations. The advantage is that the energetics of the alloy system can typically be described by a cluster expansion containing a finite number of terms T. The set of correlations that need to be considered is then simply $[-1,1]^T$, a set of perfectly manageable size. All other correlations are irrelevant, since they do not affect the energy. This transformation also allows us to represent the ground state problem as an optimization problem with continuous variables instead of discrete ones.

Formulated in this fashion the problem appears almost trivial: we have to minimize a linear function of a finite number of continuous variables. The complication comes from the fact that while any physical structure is associated with a vector of correlations $\vec{\rho} = (\rho_{\alpha_1} \dots \rho_{\alpha_T})$, not all correlations vectors correspond to a physical structure. To illustrate the problem, consider a triangular lattice with a cluster expansion that includes only a point and a nearest neighbor pair correlation. On the triangular lattice, it is impossible to construct a structure having a nearest neighbor correlation of -1. A correlation of -1 implies that every nearest neighbor pair has to connect dissimilar atoms, which is impossible on a triangular lattice. This phenomenon is referred to as "frustration".

The fundamental problem is to describe the subset S of $[-1,1]^T$ that correspond to physical structures and minimize the energy over that subset. The subset S is called the *configurational polytope* and any correlation vector in S is said to be constructible because a structure having those correlations can be constructed. An important observation made by Allen and Cahn [6] is that the configurational polytope is defined by a set of linear inequalities. Before describing how these inequalities are determined, let us note that an apparently intractable problem has been converted into the surprisingly simple task of minimizing a *linear* function over a set defined by *linear* inequalities. This very standard problem can be solved through linear programming techniques [31]. The essential idea is that a linear function defined on a convex[2] polytope necessarily reaches its minimum at one of the vertices[3] of the polytope. Thus, the ground state problem reduces to finding the vertex of the configurational polytope

[2]The convexity of S can be shown be considering any two points $\vec{\sigma}_1$ and $\vec{\sigma}_2$ in S and noting that a structure having correlation vector lying on the line joining $\vec{\sigma}_1$ and $\vec{\sigma}_2$ can be constructed by forming a phase-separated mixture of the structures associated with $\vec{\sigma}_1$ and $\vec{\sigma}_2$. The convexity of S drastically simplifies the search for the minimum.

[3]Even if the minimum were reached at more than one point at the same time, at least one of these points would be a vertex.

that minimizes energy.

To obtain the set of structures the minimize energy as a function of composition, all is needed is to add a chemical potential term to the cluster expansion: $\mu\rho_\pi$ where π is the point cluster.[4] By scanning every possible value of the chemical potential μ, every ground state is found in turn.

5.2.2 Determination of the configurational polytope

We now return to the problem of determining the configurational polytope. The essential idea is that, while it is difficult to prove that a given correlation vector $\vec{\rho}$ corresponds to an actual structure on an *infinite* lattice, it is very simple to prove that $\vec{\rho}$ corresponds to an actual structure defined on a *finite* cluster of lattice sites $\alpha_{\max}$. In other words, we no longer try to construct periodic structures on an infinite lattice, but only to construct configurations on a finite cluster $\alpha_{\max}$. We consider as constructible (on a cluster) any correlation vectors $\vec{\rho}$ that can be obtained by a weighted average of the correlations of a set of configurations defined on $\alpha_{\max}$. Note that $\alpha_{\max}$ must be chosen so that it includes all the clusters $\alpha_1 \ldots \alpha_T$ considered in the cluster expansion.

To simplify the notation, we introduce the following convention. The statement $\sum_{\alpha \hat{\subset} \alpha_{\max}}$ denotes any subcluster α of $\alpha_{\max}$, excluding clusters that are equivalent by a symmetry operation of the parent lattice to clusters already considered. It is a shorthand notation for the set of all symetrically distinct subsets of $\alpha_{\min}$. As shown in [41], the configurational polytope derived from a finite cluster $\alpha_{\max}$ is given by the set of vectors $(\rho_\alpha)_{\alpha \hat{\subset} \alpha_{\max}}$ such that, for any spin configuration σ on $\alpha_{\max}$,

$$\sum_{\alpha \hat{\subset} \alpha_{\max}} V_{\sigma\alpha}\rho_\alpha \geq 0 \tag{5.1}$$

where

$$V_{\sigma\alpha} = \sum_{\gamma \in \Omega(\alpha, \alpha_{\max})} \prod_{i \in \gamma} \sigma_i$$

where $\Omega(\alpha, \alpha_{\max})$ is the set of all clusters symmetrically equivalent to α that lie in $\alpha_{\max}$ and σ_i is the spin on site i when cluster α has configuration σ.

We briefly outline the proof of this expression. Let us momentarily ignore symmetry of the lattice. By definition, the correlation associated with cluster γ for configuration σ on $\alpha_{\max}$ is simply given by $\prod_{i \in \gamma} \sigma_i$, which denote by $v_{\gamma\sigma}$. Let us assign a probability $P_\sigma \geq 0$ to each configuration σ. The average correlation of this mixture of configurations is then given by

$$\rho_\gamma = \sum_\sigma v_{\gamma\sigma} P_\sigma. \tag{5.2}$$

[4] If there are more than one symmetrically distinct point clusters: $\mu \sum_i m_{\pi_i} \rho_{\pi_i}$.

The correlation vector $(\rho_\gamma)_{\gamma\in\alpha_{\max}}$ is constructible on $\alpha_{\max}$ because it is the average correlation of a set of actual configurations on $\alpha_{\max}$. It can be shown that the matrix $v_{\gamma\sigma}$ is orthogonal[5] and it follows that Equation (5.2) can be directly inverted:

$$P_\gamma = 2^{-|\alpha_{\max}|}\sum_\gamma v_{\gamma\sigma}\rho_\sigma$$

where $|\alpha_{\max}|$ denotes the number of sites included in $\alpha_{\max}$. Collecting terms that are equivalent by symmetry then yields Equation (5.1).

Constructibility on an infinite lattice implies constructibility on a cluster. To see this, consider every possible cluster on the infinite lattice that is symmetrically equivalent to $\alpha_{\max}$. For a given structure defined on the infinite lattice, the structure is clearly constructible on each image of the cluster $\alpha_{\max}$. Also, the overall average of the correlation on each image of the cluster $\alpha_{\max}$ is the same as the average correlation of the whole infinite lattice.

Since constructibility on an infinite lattice implies constructibility on a cluster, the configurational polytope obtained from a finite cluster $S_{\alpha,\max}$ necessarily contains the true configurational polytope S. This observation is extremely important because it allows us to *prove* that a given structure is a ground state without considering every possible structure. We first find the minimum $\vec{\rho}^*$ of the energy over $S_{\alpha,\max}$, an operation that can be performed in a finite amount of time (since $\alpha_{\max}$ is finite). Now, suppose that we are able to find a structure on the infinite lattice that has the correlation vector $\vec{\rho}^*$. The precise procedure to accomplish this is described in Appendix B but for now let us simply say that this operation takes only a modest amount of time if the structure we are interested in happens to have a small unit cell.

It can then be shown that these two observations imply that $\vec{\rho}^*$ is the correlation vector of a ground state on the lattice. On the one hand,

$$\begin{aligned} E(\vec{\rho}^*) &= \min_{\vec{\rho}\in S_{\alpha,\max}} E(\vec{\rho}) \\ &\leq \min_{\vec{\rho}\in S} E(\vec{\rho}) \text{ since } S\subset S_{\alpha,\max}. \end{aligned}$$

On the other hand, since $\vec{\rho}^*$ is constructible, $\vec{\rho}^*\in S$ and

$$E(\vec{\rho}^*) \geq \min_{\vec{\rho}\in S} E(\vec{\rho}).$$

It follows that $E(\vec{\rho}^*) = \min_{\vec{\rho}\in S} E(\vec{\rho})$, *i.e.*, that $\vec{\rho}^*$ is associated with a ground state. Of course, this nice result relies on the assumption that $\vec{\rho}^*$ is constructible. What can we do if $\vec{\rho}^*$ is not constructible? We can try to find the configurational polytope with a larger cluster $\alpha'_{\max}$ that includes the previous cluster $\alpha_{\max}$. This inclusion guarantees that the new polytope is smaller. As the size of $\alpha_{\max}$ goes to infinity, $S_{\alpha,\max}$ converge to S_α and the resulting $\vec{\rho}^*$ would be necessarily constructible. The advantage of the method lies in the observation that the true ground states of an

[5] up to a multiplicative constant

alloy system can usually be obtained with a finite cluster $\alpha_{\max}$.

The truly remarkable property of this formalism is that a problem that appears to require an infinite amount of time (scanning through every structure) is reduced to a procedure that requires a finite amount of time, and yet provides a decisive answer.

5.3 Redefinition of the Problem

The traditional method of solving this problem has essentially one limitation. Although we know exactly how large the maximal cluster needs to be to capture the effect of all interactions, it is impossible to predict the cluster size needed to capture frustration effects. Worse still, frustration effects might have a much longer range than the interaction. This requires the use of a rather large maximal cluster. Since the complexity of the problem increases exponentially with cluster size, the problem rapidly becomes intractable.

To tackle this problem, we choose the maximal cluster size only on the basis of interaction range and deal with frustration effects when they manifest themselves during the construction process.

5.3.1 Definitions

We shall call the cluster which is just large enough to capture all interactions the *low* cluster. All entities pertaining to this cluster will be prefixed by the word "*low*":

Low configuration A particular disposition of atom ($\circ$ or $\bullet$) on all sites of the low cluster. A configuration is represented by the cluster containing all sites hosting an atom of type $\bullet$.

Low correlation Any correlation $\langle \sigma_\alpha \rangle$ where α is a subcluster of the low cluster.

Low inequality Inequality expressible as a linear combination of low correlations.

Elementary low inequality Inequality stating that one particular low configuration has a positive probability of occurrence.

Low configurational polyhedron Polyhedron defined by the set of all elementary low inequalities.

Low valid inequality low inequality defining a volume which encloses the low configurational polyhedron. That is, a low valid inequality is satisfied for all point in the low configurational polyhedron.

The *low* configurational polyhedron, which is defined in a rather low dimensional space, is assumed to be completely known.

In order to detect frustration effects, we would need a larger cluster. As we have done for the low cluster, we shall prefix all entities associated with this larger cluster with the word "*high*". The *high* configurational polyhedron produced by this *high* cluster lives in a rather high dimensional space and it is impossible to know all its faces. It is nevertheless conceptually useful to know the existence of this polyhedron.

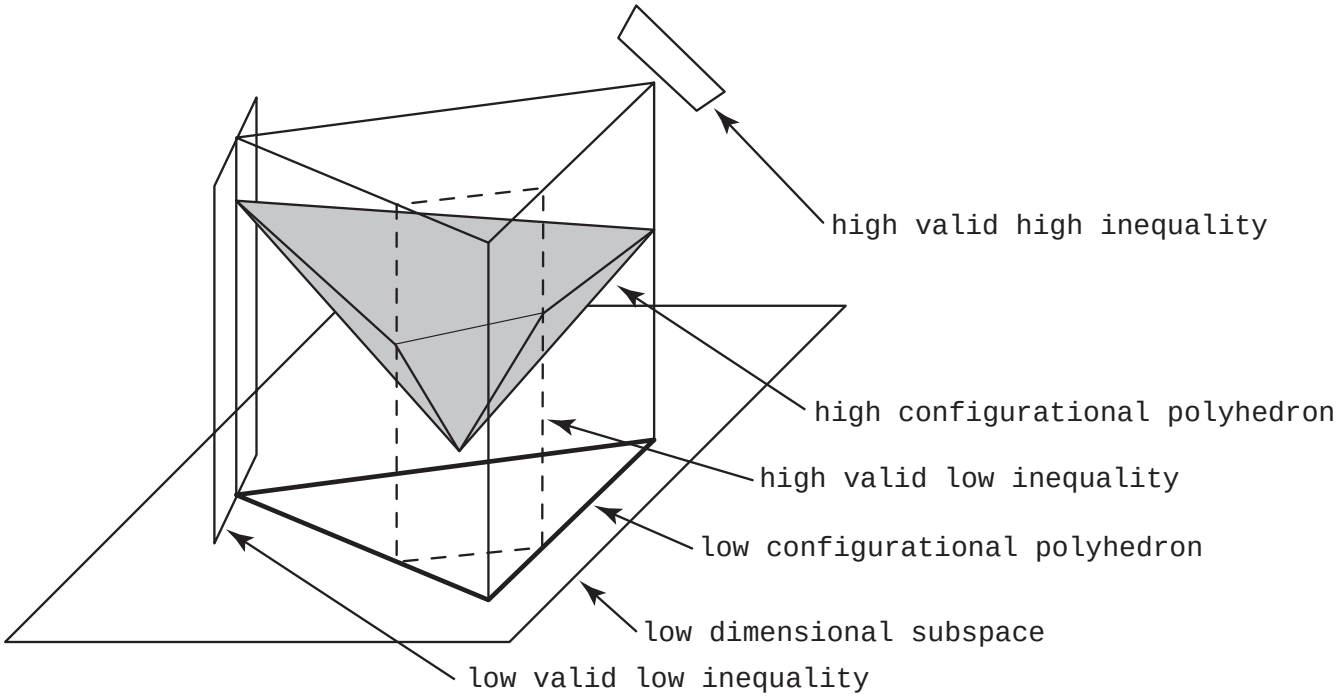


Figure 5-1: Relationship Between High-Dimensional and Low-Dimensional Quantities.

5.3.2 Properties

It is interesting to point out the relations between the "low" the "high" entities previously defined. Suppose that we embed the low dimensional space in the high dimensional one (see Figure 5-1). That is, we view the low dimensional space as an hyperplane in the high dimensional one. In this context a low inequality is just a particular type of high inequality whose normal[6] is in the low dimensional subspace (*i.e.*, the inequality is "perpendicular" to our low dimensional subspace). Similarly, any low correlation is also a high correlation.

It is also important to note that the probability of occurrence of a low configuration is equal to the sum of probability of occurrence of all high configuration containing the given low configuration on a subcluster. For example, if the low cluster is a pair and the high cluster a triplet, we may write:

$$\rho\left(\begin{smallmatrix}\circ & \bullet & \end{smallmatrix}\right) = \rho\left(\begin{smallmatrix}\circ & \bullet & \circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\circ & \bullet & \bullet\end{smallmatrix}\right)$$

where $\rho(\gamma)$ denotes the probability of occurrence of configuration γ. When the cluster α on which the configuration is defined is ambiguous we will use the notation $\rho_\alpha(\gamma)$ instead.

We shall call this property, the *expansion summation* and in general it is expressed as following. With this convention a low configuration γ on the low cluster α (*i.e.*,

[6]normal to the plane defined by changing the inequality to an equality

$\gamma \subset \alpha$) can be written as a sum of configuration on the high cluster β:

$$\rho_\alpha(\gamma) = \sum_{\delta \in \text{ expand}(\gamma,\alpha;\beta)} \rho_\beta(\delta) \tag{5.3}$$

where expand $(\gamma, \alpha; \beta)$ is the *expansion* of configuration γ on cluster α onto cluster β, defined as

$$\text{expand}\,(\gamma, \alpha; \beta) = \{\delta \subset \beta \text{ such that } (\delta \cap \alpha) = \gamma\}. \tag{5.4}$$

A consequence of the expansion summation property is that the high dimensional polyhedron is always contained within the low dimensional polyhedron. (Note that, in the high dimensional space, the low dimensional "polyhedron" is in fact a prism of infinite length whose axis is perpendicular to the low dimensional subspace.) It follows that a low valid inequality is always high valid. But the converse is not true: first, an arbitrary high inequality is generally not a low inequality. But even if we choose a high valid inequality which is a low inequality it might well intersect the low dimensional polyhedron.

5.3.3 Back to Our Problem

In fact, the high valid low inequalities which intersect the low configuration polytope are exactly the ones which we are interested in. Indeed, we seek low inequalities because we do not want to work in the high dimensional space. We also want inequalities which intersect the low configurational polyhedron because these are the ones which give us new information about how to alter the shape of the low dimensional polyhedron to take new frustration effects into account. Finally, we want high valid inequalities because otherwise they would intersect the high dimensional polyhedron and would even exclude constructible structures.

We can now restate our problem more precisely. Suppose that, in our attempt to find a ground state using the low configurational polyhedron, we obtain a vertex v (in the low dimensional space) which seems to be non constructible. *We would thus like to find all low inequalities which are high valid and which exclude vertex v.* We then add these extra constraints to our linear programming problem and find a new candidate for a ground state. We may have to repeat the process again if this new vertex happens to be also non constructible.

Is it also possible that the initial vertex v was actually constructible but had a very large unit cell. In that case, our method should indicate this failing to generate any new inequality.

5.3.4 General Principles

Let us split the problem in two parts.

We first need a convenient way to express any low inequality. In the next section we will see how to express any low (or high) inequality as a linear combination of

elementary low (or high) inequality. This will enable us to easily select only the inequalities which exclude the vertex of interest and give us way to measure the tightness of an inequality.

We then need to verify if a given low inequality is high valid. To do so, we will view any low inequality as special type of high inequality. This high inequality can then be expressed as a linear combination of elementary high inequality. We will see in subsequent sections, that when inequalities are expressed in this form, their high validity can be checked my merely verifying if all weight of the linear combination are positive. What makes this method tractable is that it is possible to express low inequality as combination of elementary high inequality without having a complete knowledge of the high dimensional polyhedron.

The last section of this report explains how to implement the suggested method.

5.4 Generating Candidate Inequalities

Before delving into the problem of generating candidate inequalities let us first review a few properties of linear combinations of linear inequalities.

5.4.1 Linear Combinations of Inequalities

Without loss of generality, all linear equalities can be written in the form

$$1 + \sum_{i=1}^{n} a_i x_i = 0, \tag{5.5}$$

where the x_i are the correlations. There is a one-to-one correspondence between any equality of the form of Equation (5.5) in *primal space* and a point $[a_1 \dots a_n]$ in *dual space*.

Consider a linear combination of a collection of equalities:

$$\sum_j \lambda_j \left(1 + \sum_{i=1}^{n} a_{ji} x_i \right) = 0.$$

For the result to be of the form of Equation (5.5), we must require that

$$\sum_j \lambda_j = 1.$$

Such a combination is called an affine combination. In dual space, the locus of all possible affine combination of a set of p (distinct) points is the $p-1$-dimensional hyperplane containing those points.

Now consider inequalities of the form

$$1 + \sum_{i=1}^{n} a_i x_i \geq 0. \tag{5.6}$$

Note that we restrict ourselves to inequalities which define a region containing the origin. (This is not a problem since, in our case, the origin corresponds to a disordered state with having equal amount of each atom, which is always a possible structure.) This restriction permits a one-to-one correspondence between any inequality of the form Equation (5.6) and a vector $[a_1 \dots a_n]$ in dual space.

Performing a linear combination of inequalities requires that $\lambda_j \geq 0$. This restriction arises from that fact that an inequality changes direction when it is multiplied by a negative number. Adding inequalities of different directions is meaningless. An affine combination having only positive weight is called a convex combination. In dual space, the locus of all convex combinations of a set of p (distinct) points is the convex hull of those points.

Although affine combinations of inequalities do not make sense when some weight are negative, it is still convenient to define an "affine combination" in the following way. An affine combination of a set of inequalities $\{Q_j \geq 0\}$ is defined as

$$\sum_j \lambda_j Q_j \geq 0 \text{ with } \sum_j \lambda_j = 1.$$

That is, we perform the affine combination of the left hand sides only and state that the resulting expression is greater that zero.

Using this definition, we can easily express an important result. Since the result holds for both the low an the high dimensional spaces, we omit the distinctive prefix.

Proposition 1 Any *inequality (valid or not) can be expressed as an "affine combination" of elementary inequalities.*

Sketch of proof: *In dual space, the set of all elementary inequalities is mapped to a set of points. For these points to be able to generate the whole dual space by affine combinations they must not lie on a hyperplane of dimension less than the dimension of the whole dual space.*

If these points were lying on such an hyperplane the configurational polyhedron in dual space would have zero volume. Thus, in primal space the configurational polyhedron would have infinite volume. Yet the configurational polyhedron is always bounded. ■

5.4.2 Excluding a Given Vertex

In the previous section, we described a way to generate all possible low inequalities from elementary low inequalities. Let us now turn to a more specific problem: find an expression for all low inequalities which exclude a given vertex.

A vertex v is uniquely defined by the set of faces going through it. These faces are elementary inequalities which are satisfied as equalities at vertex v. Let us de-

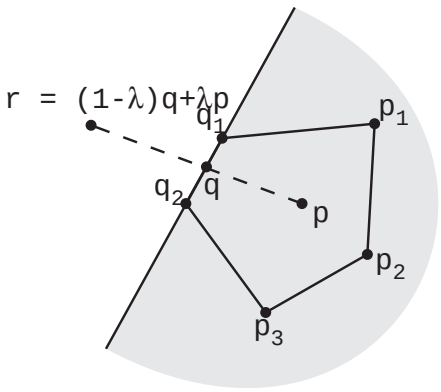


Figure 5-2: Illustration Accompanying Proposition 2.

notes these equalities by $Q_i = 0, i = 1 \ldots n$ (n is always greater or equal than the dimensionality of the space).

It is always possible to find at least another face on configurational polyhedron which does not go through v. Otherwise, the polyhedron would not be bounded. The inequalities defining such faces are satisfied as a strict inequality at v. We denote them by $P_i > 0, i = 1 \ldots m$.

We can show the following.

Proposition 2 *One can express all inequalities which exclude vertex v by combinations of the form:*

$$(1 - \lambda)Q + \lambda P \geq 0, \tag{5.7}$$

where Q is any affine combination of the Q_i, P is a convex combination of the P_i and $\lambda < 0$.

Sketch of proof: *(See Figure 5-2.) If a linear inequality defining a region R excludes a vertex v (in primal space) it means that $v \notin R$. In dual space, this fact translates as $r \notin V$, where r is the point representing region R and V is the region represented by point v.*

All the Q_i are equal to zero at vertex v, which means that they are represented, in dual space, by points q_i located on the boundary of V. The set of all affine combination of the q_i defines the boundary of V in dual space. (Using the fact that the polyhedron is bounded, one could show that the whole boundary of V is generated and not only a subset of it).

Since $P_i > 0$ at vertex v, any convex combination P of the P_i is also strictly greater than zero at vertex v. Now consider the point p in dual space, which represents inequality P in primal space. Point p belongs to V since $P > 0$ at v. Any affine combination of a point q on the boundary of V and p will yield a point $r \notin V$ if p is given a negative weight ($\lambda < 0$). ■

We can also perform the following simplification.

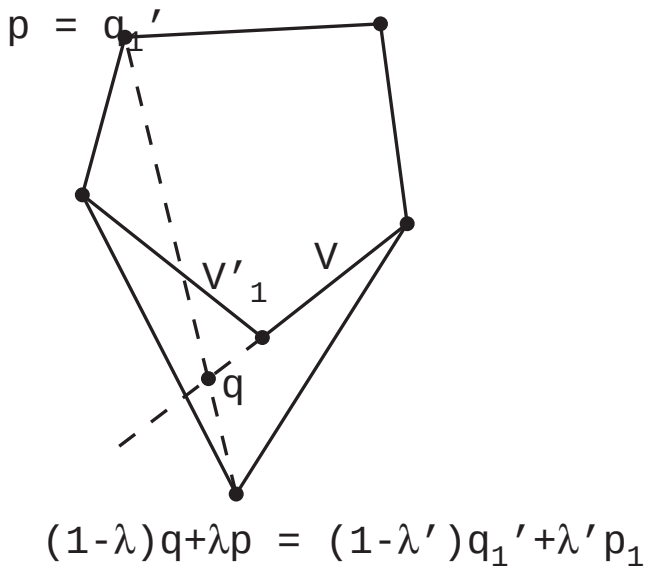


Figure 5-3: Illustration Accompanying Proposition 3.

Proposition 3 *In Proposition 2 we can restrict Q to be a convex combination instead of an affine combination.*
Sketch of proof: *We first note that whenever Q is not a convex combination, $(1-\lambda)Q + \lambda P \geq 0$ is violated at one or more other vertices v'_i (see Figure 5-3). Defining quantities Q'_i and P'_i for vertices v'_i analogously to Q and P for vertex v, we can prove that for at least one of those vertices, Q'_i is a convex combination. Hence we could have generated this candidate inequality by attempting to exclude a vertex v'_i using a convex combination for Q'_i instead of excluding vertex v using a affine combination for Q.* ■

Hence, we now have a way to generate all candidate inequalities. By using combination of elementary low inequalities, we automatically generate low inequalities. By selecting the appropriate signs in the combination, we can select the inequalities which exclude the appropriate vertex. We give positive (≥ 0) weights to inequalities which are satisfied as equalities at vertex v and negative weights (≤ 0) to the others.

5.4.3 Inequality Tightness

It would be useful to be able to determine the tightness of the inequalities we generate (*i.e.*, how much they cut out of the configurational polyhedron). That way, we could first consider the tightest low inequality and if they turn out to be high valid, we immediately add them to the linear programming problem. If this new inequality is sufficient to generate a vertex which is optimum and constructible, we can stop the process without having to generate any new inequality. On the other hand, if we generate inequalities in random order of tightness, it is unlikely that the first few generated inequalities will be sufficient to produce an optimum constructible vertex.

An inequality $A \geq 0$ is tighter than an inequality $B \geq 0$ if $A \geq 0 \Rightarrow B \geq 0$. There

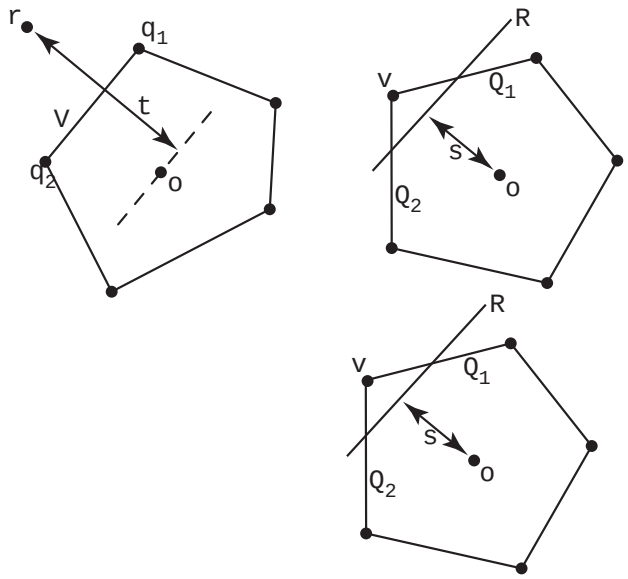


Figure 5-4: The Concept of Tightness of an Inequality (see text).

is no way to define an absolute measure of tightness: A and B can be such that no logical relation exist between them (for example, take $x + y \geq 0$ and $x - y \geq 0$). We will however define a measure of tightness which is simple to compute and agrees with the strict definition of tightness when comparison is possible.

Consider a vertex v which we try to exclude, lying at the intersection of planes $Q_i = 0$ and an inequality $R \geq 0$ whose tightness we want to evaluate (see Figure 5-4). In dual space these objects map, respectively, to a hyperplane $V = 0$ going through a set of points q_i and a point r whose "tightness" we want to evaluate.

Let us define the tightness as

$$t = \frac{r \cdot v}{\|v\|}. \tag{5.8}$$

Hence, the farther r is from plane $V = 0$, the tighter is $R \geq 0$. In primal space, this quantity also has a simple interpretation: define s as being the distance between the origin and the point of intersection of hyperplane R along the line joining the origin and vertex v:

$$s = \frac{\|v\|}{r \cdot v}.$$

We see that $t = 1/s$. In Figure 5-4, we also see that when two inequalities are parallel (*i.e.*, when their tightness can be compared), distance s lets us determine the tightest inequality. Hence t agrees with the strict definition of tightness when it applies.

We can relate the maximization of t to the value of λ in Proposition 2 by noting

that, in dual space,

$$\begin{aligned} t &= \frac{r \cdot v}{\|v\|} \\ &= \frac{((1-\lambda)q + \lambda p) \cdot v}{\|v\|} \\ &= \frac{(1-\lambda) + \lambda p \cdot v}{\|v\|} \text{ (since } p \text{ goes through plane } V,\ p \cdot v = 1) \\ &= \frac{1 + \lambda(p \cdot v - 1)}{\|v\|}. \end{aligned}$$

But $p \cdot v < 1$ because inequality $P > 0$ is satisfied at vertex v. Thus, the more negative λ is, the tighter the inequality.

5.4.4 From Rational to Integral Weights

Proposition 2 gives us an expression of all candidate inequalities in term of a affine combination of elementary inequalities. Since all inequalities of interest have rational coefficients, it follows that the weights used in the affine combination are also rational. We can always multiply the whole expression by some integer so that rational weights become integral weights.

Strictly speaking, we no longer have an affine combination but it is possible to transform any linear combination into a affine combination by simply multiplying the resulting equation by some constant so that the constant term remains 1. And multiplying an inequality by some constant does not change the region of space it represents. We will therefore not bother about the distinction between linear and affine combination unless we want to represent equalities in dual space.

This observation lets us write any candidate inequality in a more practical form:

$$\sum_{\gamma} m_\gamma \rho_\alpha(\gamma) \geq 0 \tag{5.9}$$

with $m_\gamma \in \mathrm{Z}$ or, more precisely,

$$\sum_{\gamma \in Z} m_\gamma \rho_\alpha(\gamma) - \sum_{\gamma \in \overline{Z}} |m_\gamma| \, \rho_\alpha(\gamma) \geq 0 \tag{5.10}$$

where Z is the set of all low configurations of zero probability of occurrence at vertex v while $\overline{Z}$ is the set of low configurations of non zero probability of occurrence at vertex v.

Under this point of view, maximizing our criterion of tightness implies that we must make the m_γ with $\gamma \in Z$ as small as possible and make the $|m_\gamma|$ with $\gamma \in \overline{Z}$ as large a possible, since:

$$\lambda = \frac{-\sum_{\gamma \in \overline{Z}} |m_\gamma|}{\sum_{\gamma \in Z} m_\gamma - \sum_{\gamma \in \overline{Z}} |m_\gamma|}$$

where λ is defined as in Proposition 2.

This result suggests to test inequalities in a particular order. Since λ must be strictly smaller than zero, we need at least one negative term $-\rho_\alpha(\gamma)$ with $\gamma \in \overline{Z}$. We then gradually add positive terms of the form $\rho_\alpha(\gamma)$ with $\gamma \in Z$ until a high valid inequality has been found.[7] At this point, we have at least succeeded in proving that the vertex v we have failed to construct is really non constructible.

After adding our new inequality to our linear programming problem, we check if we obtain an optimum constructible vertex. If not, we try to improve our inequality by adding another negative term and gradually adding positive terms again, until a high valid inequality has been found. We then repeat the process described in this paragraph over again until an optimum constructible vertex is found.

This algorithm generates high valid inequality in order of decreasing tightness in the following sense. Start two copies of a program which attempt to find a high valid inequality. The only difference between the two programs is that, at each step, they do not choose exactly the same terms $\rho_\alpha(\gamma)$ to add or subtract from the inequality. The first program to find a valid inequality will have found a tighter inequality than the first one generated by the other program a moment later. A similar property holds for subsequent inequalities. However there is no relation between the tightness of the n-th inequality generated by one program and the m-th inequality generated by the other program if $n \neq m$.

5.5 Testing Inequality Validity

We now describe a way to detect which candidate inequalities are high valid.

5.5.1 Convex Combination and Validity

The concept of convex combination is directly related to the concept of validity of an inequality. We will omit the prefix "low" or "high" since the following result is general.

Proposition 4 *An inequality is valid if and only if it is expressible as a convex combinations of elementary inequality.*

Proof: *To show that a convex combinations of elementary inequalities is valid we note that elementary inequalities are satisfied for any point of the configurational polyhedron. A sum of satisfied inequality always gives a satisfied inequality. Multiplying each inequality in the sum by a positive weight preserves this property.*

To show the converse, consider the half-hyperplane $A = \{\mathbf{x} \in \mathrm{R}^n$ *such that* $1 + \mathbf{a} \cdot \mathbf{x} \geq 0\}$ *and some point* $\mathbf{b} \in \mathrm{R}^n$ *(see Figure 5-5). In dual space, the point* $\mathbf{b}$ *is mapped to the half-hyperplane* $B = \{\mathbf{y} \in \mathrm{R}^n$ *such that* $1 + \mathbf{y} \cdot \mathbf{b} \geq 0\}$ *while the region* A *is mapped to a point* $\mathbf{a} \in \mathrm{R}^n$*. This shows that* $\mathbf{b} \in A$ *if and only if* $\mathbf{a} \in B$*.*

Now consider some valid inequality, represented by a half-hyperplane A*. Then all vertices* v_i *of the configurational polyhedron (in primal space) belong to* A*. In dual*

[7] How to do so will be explained in the next section.

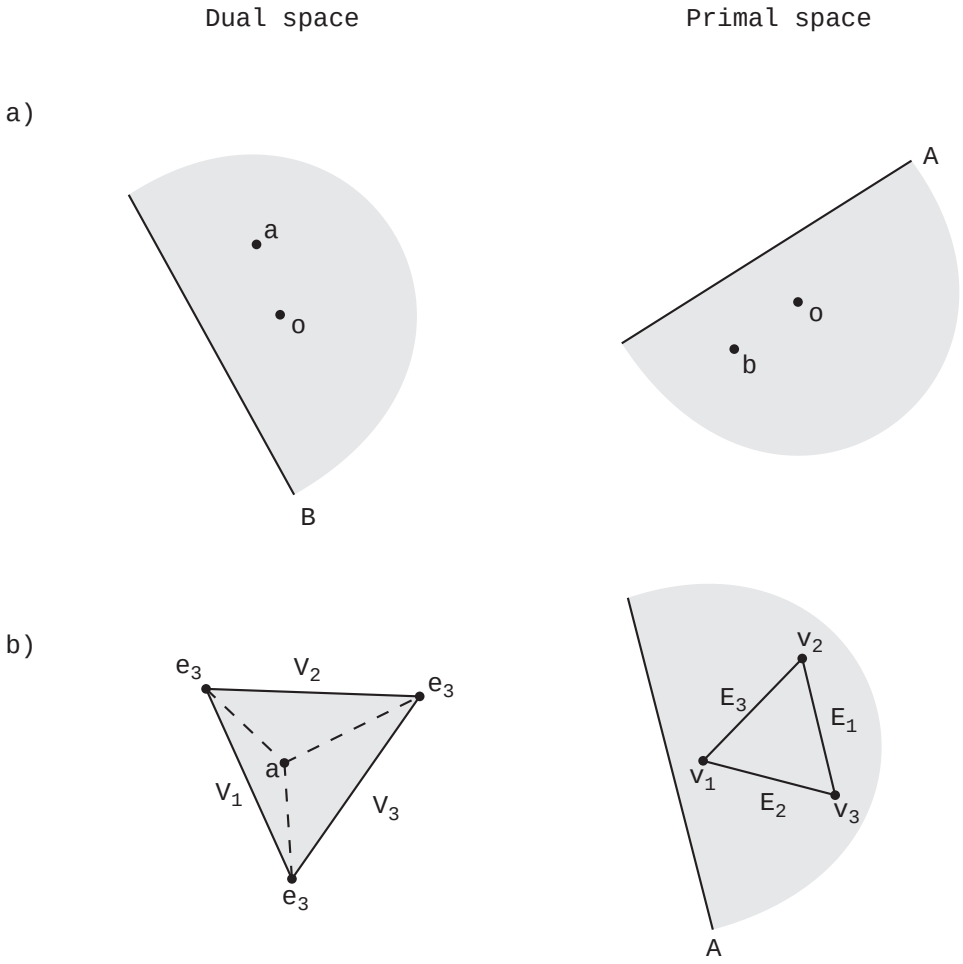


Figure 5-5: Illustration Accompanying Proposition 4.

space, all vertices become half-hyperplanes V_i while A becomes a point a which belongs to all V_i. The intersection of all V_i defines a convex polygon P. Since P is convex, $a \in P$ can be expressed as a convex combination of the vertices of P. These vertices correspond to the edges of the configurational polyhedron in primal space. Hence, any valid inequality can be expressed as a convex combination of elementary inequality. ■

Note that, in the statement of this proposition, the use of "is expressible as" instead of "is" is important. It is not because one has found a way to express an inequality as an affine combination with some negative weights that the inequality is necessarily not valid. One might find another expression which involves only positive weights. The representation of an inequality by a affine combination is generally not unique. On the contrary, as soon as one finds an expression for an inequality which is a convex combination one knows that this inequality *must* be valid.

However, in the case where the polyhedron is a simplex, there is only one way to express a given inequality as a convex combination of elementary inequalities. In this case, we can rewrite our proposition in a simpler form.

Proposition 5 *If the dimensional polytope is a simplex, then an inequality is valid if and only if it is a convex combinations of elementary inequality.*

5.5.2 Converting Low Inequalities to High Inequalities

Up to now, we have seen that we can express a low valid inequality as a convex combination of elementary low inequalities and that we can express high valid inequality a convex combination of elementary high inequalities.

But what we really need is to test the high validity of a low inequality. We now recall that a low inequality is just a particular case of a high inequality which happens to be perpendicular to the low dimensional subspace. Hence, we can also express a low inequality as an affine combination of elementary high inequalities. If this affine combination turns out to be a convex combination, this shows that our low inequality is in fact high valid.

How to perform such an operation? By Equation (5.9), we already have an expression of all candidate inequalities as an affine combination of elementary low inequalities $\rho_\alpha(\gamma) \geq 0$:

$$\sum_\gamma m_\gamma \rho_\alpha(\gamma) \geq 0. \tag{5.11}$$

A simple way to proceed is then to express each $\rho_\alpha(\gamma)$ term as a affine combination of elementary high inequalities ($\rho_\beta(\delta) \geq 0$) using the expansion summation property (Equation (5.3)):

$$\rho_\alpha(\gamma) = \sum_{\delta \in \text{ expand}(\gamma,\alpha;\beta)} \rho_\beta(\delta). \tag{5.12}$$

Equation (5.12) can be written in a more convenient way:

$$\rho_\alpha(\gamma) = \sum_\delta b_{\gamma\delta}\rho_\beta(\delta) \tag{5.13}$$

where

$$b_{\gamma\delta} = \begin{cases} 1 & \text{if } \delta \in \text{ expand}\,(\gamma, \alpha; \beta) \\ 0 & \text{otherwise .} \end{cases}$$

Combining Equation (5.11) and Equation (5.13) yields

$$\sum_\delta \left(\sum_\gamma m_\gamma b_{\gamma\delta} \right) \rho_\beta(\delta) \geq 0. \tag{5.14}$$

If $\sum_\gamma m_\gamma b_{\gamma\delta} > 0$ for all δ we have a convex combination and Equation (5.14) defines a low inequality which high valid, as desired.

Unfortunately, things are rarely so simple, as we will see in the next sections dealing with symmetry, unicity and practical considerations.

5.5.3 Using Symmetry

We first introduce two convenient definitions.

Definition 1 *Two clusters (or two configurations) γ_1 and γ_2 are α-equal (denoted $\gamma_1 \overset{\alpha}{=} \gamma_2$) if and only if there is a symmetry operation s such that $s(\alpha) = \alpha$ and $s(\gamma_1) = \gamma_2$.*

Definition 2 *The infinite cluster describing the whole lattice is denoted by ∞. Hence equivalence under a lattice symmetry operation can be denoted by $\overset{\infty}{=}$.*

The symmetry of the lattice requires that two correlation $\langle\sigma_\alpha\rangle$ and $\langle\sigma_\beta\rangle$ such that $\alpha \overset{\infty}{=} \beta$ must be equal. A consequence of this is that two configurations γ_1 and γ_2 on a cluster α must have the same probability of occurrence if $\gamma_1 \overset{\alpha}{=} \gamma_2$. It also follows that the inequalities $\rho_\alpha(\gamma_1) \geq 0$ and $\rho_\alpha(\gamma_2) \geq 0$ define exactly the same face of the configurational polyhedron.

Hence we must regroup the terms of Equation (5.13) which contain probabilities which should be equal due to symmetry. For example

$$\rho\,(\circ) = \rho\left(\begin{smallmatrix} \circ \\ \circ\,\circ \end{smallmatrix}\right) + \rho\left(\begin{smallmatrix} \circ \\ \circ\,\bullet \end{smallmatrix}\right) + \rho\left(\begin{smallmatrix} \bullet \\ \circ\,\circ \end{smallmatrix}\right) + \rho\left(\begin{smallmatrix} \bullet \\ \circ\,\bullet \end{smallmatrix}\right)$$

must be written as

$$\rho\,(\circ) = \rho\left(\begin{smallmatrix} \circ \\ \circ\,\circ \end{smallmatrix}\right) + 2\rho\left(\begin{smallmatrix} \circ \\ \circ\,\bullet \end{smallmatrix}\right) + \rho\left(\begin{smallmatrix} \bullet \\ \circ\,\bullet \end{smallmatrix}\right).$$

In general, we have:

$$\rho_\alpha(\gamma) = \sum_{\delta \in S_\beta} c_{\gamma\delta} \rho_\beta(\delta). \tag{5.15}$$

The set S_β is obtained taking the set of all configurations on β and successively removing configurations which are β-equal to another one in the set until all redundant configurations have been removed. The array $c_{\gamma\delta}$ gives the number of configurations in the set expand $(\gamma, \alpha; \beta)$ which are β-equal to δ.

Combining Equation (5.15) with Equation (5.11) yield the desired result:

$$\sum_{\delta \in S_\beta} \left(\sum_\gamma m_\gamma c_{\gamma\delta} \right) \rho_\beta(\delta) \geq 0. \tag{5.16}$$

If $\sum_\gamma m_\gamma c_{\gamma\delta} > 0$ for all $\delta \in S_\beta$ we have a convex combination and Equation (5.16) defines a low inequality which high valid!

5.5.4 A Simple Example

The theory presented above still cannot cover all cases in a straightforward way. But before going on, let us illustrate the procedure with a simple example in which our theory applies.

Take a triangular lattice with a pair cluster as the low cluster. The only non constructible vertex v of the low configurational polytope is characterized by

$$\rho\left(\begin{smallmatrix}\bullet\\ \circ\end{smallmatrix}\right) > 0 \tag{5.17}$$
$$\rho\left(\begin{smallmatrix}\circ\\ \circ\end{smallmatrix}\right) = 0 \tag{5.18}$$
$$\rho\left(\begin{smallmatrix}\bullet\\ \bullet\end{smallmatrix}\right) = 0. \tag{5.19}$$

Following Equation (5.10) we perform an affine combination of the form

$$m_1 \rho\left(\begin{smallmatrix}\circ\\ \circ\end{smallmatrix}\right) + m_2 \rho\left(\begin{smallmatrix}\bullet\\ \bullet\end{smallmatrix}\right) - |m_3| \, \rho\left(\begin{smallmatrix}\bullet\\ \circ\end{smallmatrix}\right) \geq 0 \tag{5.20}$$

We can expand each low configuration to a high cluster:

$$\begin{aligned}
\rho\left(\begin{smallmatrix}\bullet\\ \circ\end{smallmatrix}\right) &= \rho\left(\begin{smallmatrix}\bullet & \circ\\ \circ & \end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet & \bullet\\ \circ & \end{smallmatrix}\right) \\
\rho\left(\begin{smallmatrix}\circ\\ \circ\end{smallmatrix}\right) &= \rho\left(\begin{smallmatrix}\circ & \circ\\ \circ & \end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\circ & \bullet\\ \circ & \end{smallmatrix}\right) \\
\rho\left(\begin{smallmatrix}\bullet\\ \bullet\end{smallmatrix}\right) &= \rho\left(\begin{smallmatrix}\bullet & \circ\\ \bullet & \end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet & \bullet\\ \bullet & \end{smallmatrix}\right)
\end{aligned}$$

and rewrite these equations in terms of configurations which are equivalent by sym-

metry:

$$\rho\left(\begin{smallmatrix}\circ\\ \circ\bullet\end{smallmatrix}\right) = \rho\left(\begin{smallmatrix}\bullet\\ \circ\circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \circ\bullet\end{smallmatrix}\right) \quad (5.21)$$

$$\rho\left(\begin{smallmatrix}\circ\\ \circ\circ\end{smallmatrix}\right) = \rho\left(\begin{smallmatrix}\circ\\ \circ\circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \circ\circ\end{smallmatrix}\right) \quad (5.22)$$

$$\rho\left(\begin{smallmatrix}\bullet\\ \bullet\bullet\end{smallmatrix}\right) = \rho\left(\begin{smallmatrix}\bullet\\ \circ\bullet\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \bullet\bullet\end{smallmatrix}\right) \quad (5.23)$$

Substituting into Equation (5.20) yields:

$$m_1\rho\left(\begin{smallmatrix}\circ\\ \circ\circ\end{smallmatrix}\right) + m_1\rho\left(\begin{smallmatrix}\bullet\\ \circ\circ\end{smallmatrix}\right) + m_2\rho\left(\begin{smallmatrix}\bullet\\ \circ\bullet\end{smallmatrix}\right) + m_2\rho\left(\begin{smallmatrix}\bullet\\ \bullet\bullet\end{smallmatrix}\right) - |m_3|\,\rho\left(\begin{smallmatrix}\bullet\\ \circ\circ\end{smallmatrix}\right) - |m_3|\,\rho\left(\begin{smallmatrix}\bullet\\ \circ\bullet\end{smallmatrix}\right) \geq 0.$$

Using the ordering suggested in Section 5.4.4, we start with one negative term ($m_3 = -1$). We then gradually add positive terms. Let us start by setting $m_2 = 1$ and $m_1 = 0$:

$$\rho\left(\begin{smallmatrix}\bullet\\ \circ\bullet\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \bullet\bullet\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\bullet\\ \circ\circ\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\bullet\\ \circ\bullet\end{smallmatrix}\right) \geq 0.$$

or

$$\rho\left(\begin{smallmatrix}\bullet\\ \bullet\bullet\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\bullet\\ \circ\circ\end{smallmatrix}\right) \geq 0. \quad (5.24)$$

We still have a negative term, so we try adding another positive term. Let us increment m_1 by one, which adds a term of the form $\rho\left(\begin{smallmatrix}\circ\\ \circ\circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \circ\circ\end{smallmatrix}\right)$ to Equation (5.24):

$$\rho\left(\begin{smallmatrix}\circ\\ \circ\circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \circ\circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \bullet\bullet\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\bullet\\ \circ\circ\end{smallmatrix}\right) \geq 0.$$

or

$$\rho\left(\begin{smallmatrix}\circ\\ \circ\circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \bullet\bullet\end{smallmatrix}\right) \geq 0. \quad (5.25)$$

Equation (5.25) is a convex combination of elementary high inequalities and is thus high valid. It is also a low inequality because it was obtained by taking an affine combination of low inequalities:

$$\rho\left(\begin{smallmatrix}\circ\\ \circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \bullet\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\bullet\\ \circ\end{smallmatrix}\right) \geq 0.$$

Finally, it excludes v, desired.

The problem with this method is that we rely on intuition to decide which positive m_i to increment. In a practical algorithm, multiple alternative would have to be considered. But the beauty of this method is that if one runs multiple copies of this algorithm in parallel, each copy trying a different sequence of m_i to increment, the first program to find a high valid inequality will have found the tightest inequality among the first inequalities generated by all the programs.

5.5.5 Unicity

In the previous sections, we chose to express low inequalities as affine combination of elementary high inequality using the expansion summation property. When the high configuration polyhedron is a simplex, this gives us the only way to represent a given low inequality as an affine combination of elementary high inequality. In the previous example, the high configurational polyhedron was indeed a simplex.

On the other hand, when the high configurational polyhedron is *not* a simplex, we have to try different representations of the given low inequality until we find one which is a convex combination. How do these other representations look like? Part of the non unicity arises because there is more than one way to place the low cluster within the high cluster. That is, there is more than one distinct expansion summation. For example, for a point as the low cluster and a linear triplet as the high cluster, we may write $\rho(\circ)$ as

$$\rho\,(\circ \quad) = \rho(\circ\,\circ\,\circ) + \rho(\circ\,\circ\,\bullet) + \rho(\circ\,\bullet\,\circ) + \rho(\circ\,\bullet\,\bullet)$$

or as

$$\rho\,(\ \circ\ ) = \rho(\circ\,\circ\,\circ) + \rho(\circ\,\circ\,\bullet) + \rho(\bullet\,\circ\,\circ) + \rho(\bullet\,\circ\,\bullet).$$

The right hand side of these two expressions cannot be made identical by any symmetry operation leaving the high cluster invariant. By performing all possible affine combinations of these two expressions we obtain all possible representation for $\rho(\circ) \geq 0$. In this simple case, this is sufficient to account for the non unicity. However, in more complex cases, other representations which do not arise from simple expansion summation have to be considered in the affine combination. How this can be done is still not clear.

So the morale of the story is that we really need our high configuration to be a simplex to be able to do anything with any certainty. How to choose a high cluster such that it produces a configurational polyhedron which is a simplex? The answer is curious: simply take the high cluster to be the infinite cluster representing the whole lattice ($\beta = \infty$):

Proposition 6 *The configurational polyhedron obtained using the whole lattice (∞) as the maximal cluster is a simplex.*
Sketch of proof: *Since the high cluster is the whole lattice, $\rho_\infty(\delta_1) = \rho_\infty(\delta_2)$ if $\delta_1 \overset{\infty}{=} \delta_2$. We already have a similar result for correlations: $\langle\sigma_{\delta_1}\rangle = \langle\sigma_{\delta_2}\rangle$ if $\delta_1 \overset{\infty}{=} \delta_2$. Hence we can again perform a one-to-one mapping between configuration and correlation ($\delta \leftrightarrow \langle\sigma_\delta\rangle$) to show that the configurational polyhedron is a simplex. (One-to-one mappings between infinite sets are usually a tricky matter, but the bijection used here is so natural that it is unlikely that any serious problems show up.)* ■

The use of an infinite high cluster also solves a problem which we have, until now, swept under the rug: How large must the high cluster be? In our simple example, a triplet turned out to be sufficient. But in general, we don't know. Using a infinite high cluster guaranties than we include all frustration effects.

This may look like a rather theoretical result since the expansion summation now contains an infinite number of terms. It turns out that we never need to write down the complete expansion summation to perform the operations we need.

5.6 Implementation

This section is divided as follows. We first present of a way to represent the infinite sums which arise when the high cluster is taken to be infinite. We then describe how to perform addition and subtraction with these representations.

5.6.1 Representing an Infinite Sum

Ideally, we would like to convert a affine combination of the form

$$\sum_{\gamma} m_\gamma \rho_\alpha(\gamma) \geq 0 \tag{5.26}$$

to an affine combination of the form

$$\sum_{\delta \in S_\infty} w_\delta \rho_\infty(\delta) \geq 0 \tag{5.27}$$

and check that $w_\delta \geq 0$. Theoretically, this operation could be performed as following. First use the expansion summation property (which gives a unique expression in this case),

$$\sum_{\gamma} m_\gamma \sum_{\delta \in \text{ expand}_{(\gamma,\alpha;\infty)}} \rho_\infty(\delta) \geq 0,$$

and then transform each configuration δ by some symmetry operation s_δ such that $s_\delta(\delta) \in S_\infty$:

$$\sum_{\gamma} m_\gamma \sum_{\delta \in \text{ expand}_{(\gamma,\alpha;\infty)}} \rho_\infty(s_\delta(\delta)) \geq 0. \tag{5.28}$$

Grouping common terms then yields the only expression of the form Equation (5.27) which is equal to Equation (5.26).

This approach is, of course, not practical: Equation (5.28) is an infinite sum and we must apply a different symmetry operation s_δ to each (infinitely big) configuration in the sum. We thus define an intermediate representation between Equation (5.26) and Equation (5.27)

$$\sum_{i=1}^{N} n_i \rho_{\alpha_i}(\gamma_i) \geq 0, \tag{5.29}$$

where $n_i = \pm 1$ and $|\alpha| \leq |\alpha_i| < \infty$ ($|\alpha|$ is the number of sites in α) and N is *finite*.[8] Note that configurations γ_i and clusters α_i can be repeated for different values of i.

It looks like we are back to using a finite high cluster, with all the problems it implies. However, there is one main difference. The sizes of the intermediate clusters α_i are not fixed and shall let them grow as needed. We thus avoid the problem of guessing *a priori* the cluster size needed to take all frustration effects into account. This will also enable us to bypass the unicity problem.[9]

5.6.2 General Principles

Our method proceeds as following. Instead of attempting to apply a different symmetry operation to each term in an infinite sum, we classify the terms in a finite number of categories and apply the same symmetry operation on all terms of a category. We then try to cancel all negative terms with identical positive terms. If we are unable to cancel all negative terms, it means than our classification is too coarse. We then split some categories in smaller subcategories and apply a different symmetry operation in each subcategory and again attempt to simplify identical terms. In the limit, if we need to apply this subdivision process an infinite number of times, we can actually apply a different symmetry operation to each term individually. Hopefully, we will not need to take this process to the limit.

Equation (5.29) is a way of separating terms in categories: applying a symmetry operation to one term $\rho_{s_i(\alpha_i)}(s_i(\gamma_i))$ is equivalent to transforming all configurations in $\text{expand}(\gamma_i, \alpha_i; \infty)$. To obtain a finer division into categories we use expansion summation to write

$$\rho_{\alpha_i}(\gamma_i) = \sum_{\delta \in \text{expand}(\gamma_i, \alpha_i; \beta_i)} \rho_{\beta_i}(\delta)$$

and then apply a different symmetry operation on each term $\rho_{\beta_i}(\delta)$. Our method consists essentially in repeating such a sequence a expansion, application of symmetry and cancellation of identical terms, until all negative terms have been removed. Indeed, if $n_i \geq 0$ in Equation (5.29), then it represent a high valid inequality. To show this, we note that using the expansion summation to expand configurations on an infinite cluster and then using symmetry operation to regroup equivalent terms does not change the signs of the terms.

5.6.3 Cancellation of Identical Terms

Let us now see precisely how this is done. We will only handle the case of a subtraction between two terms since multiple subtractions can be performed by merely combining

[8] Not all inequalities can be expressed in the form of Equation (5.29) but it turns out that those that we need can.

[9] Although, obviously, a given inequality has more than one representation of the form of Equation (5.29).

terms pairwise. We can write:

$$\rho_{\alpha_1}(\gamma_1) - \rho_{s(\alpha_2)}(s(\gamma_2)) = \sum_{\delta \in \text{expand}(\gamma_1,\alpha_1;\beta)} \rho_\beta(\delta) - \sum_{\delta \in \text{expand}(s(\gamma_2),s(\alpha_2);\beta)} \rho_\beta(\delta)$$

where $\beta = \alpha_1 \cup s(\alpha_2)$. Now, both terms represent a list of configuration probability defined on a common cluster β and it is a simple matter to check whether all terms generated by $-\rho_{s(\alpha_2)}(s(\gamma_2))$ can be cancelled by identical terms generated by $\rho_{\alpha_1}(\gamma_1)$. A few examples will clarify the procedure.

$$\begin{aligned}
\rho\left(\begin{smallmatrix}\bullet\\ \circ\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\circ\\ \circ\end{smallmatrix}\right) &= \rho\left(\begin{smallmatrix}\bullet\ \\ \circ\ \end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\ \circ\\ \ \circ\end{smallmatrix}\right) \\
&= \rho\left(\begin{smallmatrix}\bullet & \\ \circ & \circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix} & \bullet\\ \circ & \bullet\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix} & \circ\\ \circ & \circ\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix} & \bullet\\ \circ & \circ\end{smallmatrix}\right) \\
&= \rho\left(\begin{smallmatrix} & \bullet\\ \circ & \bullet\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix} & \circ\\ \circ & \circ\end{smallmatrix}\right)
\end{aligned}$$

$$\begin{aligned}
\rho(\circ) - \rho\left(\begin{smallmatrix}\bullet\\ \circ\end{smallmatrix}\right) &= \rho\left(\begin{smallmatrix}\ \\ \circ\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\bullet\\ \circ\end{smallmatrix}\right) \\
&= \rho\left(\begin{smallmatrix}\circ\\ \circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \circ\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\bullet\\ \circ\end{smallmatrix}\right) \\
&= \rho\left(\begin{smallmatrix}\circ\\ \circ\end{smallmatrix}\right)
\end{aligned}$$

These two examples illustrate two important results.

Proposition 7

$$\rho_{\alpha_1}(\gamma_1) - \rho_{\alpha_2}(\gamma_2) = \sum_{\delta \in A} \rho_\beta(\delta) - \sum_{\delta \in B} \rho_\beta(\delta)$$

where $\beta = \alpha_1 \cup \alpha_2$ *and*

$$\begin{aligned}
A &= \{\delta \subset \beta \text{ such that } (\delta \cap \alpha_1) = \gamma_1 \text{ and } (\delta \cap \alpha_2) \neq \gamma_2\} \\
B &= \{\delta \subset \beta \text{ such that } (\delta \cap \alpha_1) \neq \gamma_1 \text{ and } (\delta \cap \alpha_2) = \gamma_2\}.
\end{aligned}$$

Proof: *The only configuration* δ *that commonly belongs to expand*$(\gamma_1, \alpha_1; \beta)$ *as well as expand*$(\gamma_2, \alpha_2; \beta)$ *is the one for which* $\delta \cap \alpha_1 = \gamma_2$ *and* $\delta \cap \alpha_2 = \gamma_2$*, hence the result.* ■

This proposition essentially tells which terms survive when we can partially overlap the two configurations γ_1 and γ_2 so that the type of atom agrees at each site which belongs to both α_1 and α_2.

Proposition 8 *If* $\alpha_1 \subset \alpha_2$ *and* $\gamma_2 \cap \alpha_1 = \gamma_1$ *then*

$$\rho_{\alpha_1}(\gamma_1) - \rho_{\alpha_2}(\gamma_2) = \sum_{\delta \in A} \rho_\beta(\delta)$$

where $\beta = \alpha_1 \cup \alpha_2$ *and*

$$A = \{\delta \subset \beta \text{ such that } (\delta \cap \alpha_1) = \gamma_1 \text{ and } (\delta \cap \alpha_2) \neq \gamma_2\}$$

Proof: *Application of the Proposition 7 gives the desired result.* ■

This proposition tells us that no negative term survives whenever we can find a configuration γ_1 on a subcluster α_1 of a cluster α_2 on which configuration γ_2 is defined.

5.6.4 Simple Example

Proposition 7 appears to be just procrastinating the problem: a difference is still expressed as a difference. The only improvement is an increase of the cluster sizes. When a configuration is defined on a bigger cluster, it is more likely that Proposition 8 applies. For example, consider the expression

$$-\rho\left(\begin{smallmatrix}\circ\\ \bullet\end{smallmatrix}\right)+\rho\left(\begin{smallmatrix}\circ\\ \circ\end{smallmatrix}\right)+\rho\left(\begin{smallmatrix}\bullet\\ \bullet\end{smallmatrix}\right)\geq 0.$$

Since there is not much we can say about this expression, let us try to combine the first two terms using some symmetry operation

$$-\rho\left(\begin{smallmatrix}\circ\\ \bullet\end{smallmatrix}\right)+\rho\left(\begin{smallmatrix}\circ & \circ\end{smallmatrix}\right)+\rho\left(\begin{smallmatrix}\bullet\\ \bullet\end{smallmatrix}\right)\geq 0$$

and Proposition 7

$$-\rho\left(\begin{smallmatrix}\circ & \bullet\\ \bullet & \bullet\end{smallmatrix}\right)+\rho\left(\begin{smallmatrix}\circ & \circ\\ \circ & \end{smallmatrix}\right)+\rho\left(\begin{smallmatrix}\bullet\\ \bullet\end{smallmatrix}\right)\geq 0.$$

Again using symmetry and exchanging terms yields

$$\rho\left(\begin{smallmatrix} & \bullet\\ \bullet & \bullet\end{smallmatrix}\right)-\rho\left(\begin{smallmatrix}\circ & \bullet\\ \bullet & \bullet\end{smallmatrix}\right)+\rho\left(\begin{smallmatrix}\circ & \circ\\ \circ & \end{smallmatrix}\right)\geq 0$$

which lets us use Proposition 8 to obtain:

$$\rho\left(\begin{smallmatrix}\bullet & \bullet\\ \bullet & \bullet\end{smallmatrix}\right)+\rho\left(\begin{smallmatrix}\circ & \circ\\ \circ & \end{smallmatrix}\right)\geq 0,$$

proving that the expression represents a high valid inequality.

5.6.5 Choosing the Best Symmetry

Up to now we have used intuition to figure out which symmetry to use on the configurations before using Propositions 7 and 8 so that the maximum number of terms will cancel. But there is a rigorous way to do so. We first define a measure of how close we are to our goal of eliminating all negative terms. We will choose the symmetry operation which minimizes this measure.

The raw number of negative terms is clearly not a good measure. For example, take a cluster of n sites, in which c configurations have negative weight. If one expands these configuration on a cluster of size $m > n$, one will have $2^{m-n}c$ configurations of negative weight without having done anything. A proper measure of success M would then be the ratio of the number of configuration of negative weight[10] over the total

[10]The number of term of the form $-\rho_\alpha(\gamma)$ in the expression whose distance to the goal is to be

number of possible configuration on a cluster of that size. (Repeated configuration are counted more than once in the numerator while symmetry-related configuration are considered different in the denominator.) More precisely, in Equation (5.29) we have

$$M = \sum_{i \text{ such that } n_i<0} \frac{-n_i}{2^{|\alpha_i|}}. \tag{5.30}$$

Let us consider which choices of symmetry operation have no effect on M in the expression

$$\rho_{\alpha_1}(\gamma_1) - \rho_{s(\alpha_2)}(s(\gamma_2)).$$

- Suppose that we apply a symmetry operation s to α_2 and γ_2 so that α_1 and $s(\alpha_2)$ do not overlap. Application of Proposition 7 shows that

  $$M = \frac{2^{\alpha_1}}{2^{|\alpha_1|+|\alpha_2|}} = \frac{1}{2^{|\alpha_2|}}$$

  because no term cancels out.

- Suppose s is such that α_1 and $s(\alpha_2)$ overlap but such that $\gamma_1 \cap (\alpha_1 \cap s(\alpha_2)) \neq s(\gamma_2) \cap (\alpha_1 \cap s(\alpha_2))$ (*i.e.*, the two configurations do not match at the sites where the clusters overlap). Then application of Proposition 7 yields the same result

  $$M = \frac{2^{\alpha_1 \setminus s(\alpha_2)}}{2^{|\alpha_1 \cup s(\alpha_2)|}} = \frac{1}{2^{|\alpha_2|}}.$$

On the other hand, if s is such that α_1 and $s(\alpha_2)$ overlap and such that $\gamma_1 \cap (\alpha_1 \cap s(\alpha_2)) = s(\gamma_2) \cap (\alpha_1 \cap s(\alpha_2))$ then

$$M = \frac{2^{|\alpha_1 \setminus s(\alpha_2)|} - 1}{2^{|\alpha_1 \cup s(\alpha_2)|}}.$$

This result arises from the fact that configuration $s(\gamma_2)$ expands to a set of $2^{|\alpha_1 \setminus s(\alpha_2)|}$ configurations out of which only one is cancelled (see Proposition 7). This expression can be rewritten as:

$$M = \frac{1 - \frac{2^{|\alpha_1 \cap s(\alpha_2)|}}{2^{|\alpha_1|}}}{2^{|\alpha_2|}}.$$

At constant $|\alpha_1|$ and $|\alpha_2|$, the bigger $|\alpha_1 \cap s(\alpha_2)|$ is, the smaller is M. (It the limit case where $\alpha_1 \subset s(\alpha_2)$, $M = 0$, as expected from the conclusion of Proposition 8.) Hence, we seek to overlap cluster α_1 and $s(\alpha_2)$ as much as possible while making sure that the configurations on both clusters agree at their intersection. If the overlap is

evaluated.

complete ($\alpha_1 \subset s(\alpha_2)$), then $\rho_{\alpha_1}(\gamma_1) - \rho_{\alpha_2}(\gamma_2)$ can be expressed as a sum containing only positive terms.

This result tells us how to evaluate the "convexity" of a difference of two terms. In a practical algorithm, this procedure would be repeated over and over. First start with a positive and negative term. If all negative terms cannot be cancelled add a new positive term to combine the remaining negative terms. The process is repeated until all negative terms have been removed and a high valid inequality $Q \geq 0$ has been found.

To obtain yet tighter inequalities, if they are needed, we can add another negative term to Q, perform a few cancellation with existing positive term and then add on new positive terms if they are needed. We can even loop back one again at the beginning of this paragraph if no ground state has been found yet. This process fits nicely with the method outlined at the end of Section 5.4.4.

5.6.6 Avoiding Infinite Loops

This section is somewhat technical and can be skipped at first reading. It is included for completeness.

We have to handle a subtle question which we have ignored in the previous section: Shouldn't we try to apply Proposition 7 again and again before attempting to add new positive terms? [11] For example, after writing

$$-\rho\left(\begin{smallmatrix}\circ\\ \circ & \bullet\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\circ\\ \circ & \circ\end{smallmatrix}\right) = -\rho\left(\begin{smallmatrix}\bullet\\ \circ & \bullet\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\circ\\ \circ & \circ\end{smallmatrix}\right)$$

we could apply Proposition 7 again to obtain

$$\begin{aligned} -\rho\left(\begin{smallmatrix}\circ\\ \circ & \circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\\ \circ & \bullet\end{smallmatrix}\right) \;=\; & -\rho\left(\begin{smallmatrix}\circ\\ \circ & \circ & \bullet\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\circ\\ \circ & \bullet & \circ\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\circ\\ \circ & \circ & \circ\end{smallmatrix}\right) \\ & + \rho\left(\begin{smallmatrix}\bullet\\ \circ & \bullet & \bullet\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\circ\\ \bullet & \circ & \bullet\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\bullet\\ \bullet & \circ & \bullet\end{smallmatrix}\right). \end{aligned}$$

We can repeat the process an infinite number of times and, at each step, M decreases! (It never never goes to zero, though.)

We thus need a criterion to determine whether a subtraction of the form $\rho_{\alpha_1}(\gamma_1) - \rho_{\alpha_2}(\gamma_2)$ cannot be expressed as a sum of term. Here are cases where it is surely not:

- when $\alpha_1 \overset{\infty}{=} \alpha_2$ and $\gamma_1 \overset{\alpha_1}{\neq} s(\gamma_2)$ where s is such that $s(\alpha_2) = \alpha_1$ (because two different elementary inequality on the same cluster necessarily have different expression in the form of Equation (5.27));

- when $|\alpha_1| > |\alpha_2|$ (because there is simply not enough positive terms to cancel all negative terms).

There are lot of other cases but they are cumbersome to describe and to prove. In fact, unless really exceptional circumstances occur,[12] repeatedly applying Proposition

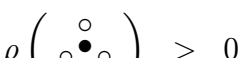
[11] Recall that adding new terms decreases the tightness of the generated inequality.

[12] The only problematic case for which a proof as not yet been found is when there exist a symmetry

7 never allows us to remove all negative term if it did not work at the first step. Thus, almost surely, the conclusion is: use Proposition 7 once and upon failure to remove all negative terms, try adding new positive terms.

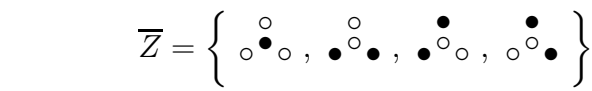
5.6.7 Optimization

In Proposition 7 we perform cancellation only one term at the time. However, there is a way to perform many cancellations in one operation. Let us first illustrate the method in the case of the honeycomb "lattice". If we use $\alpha =$  as the low cluster,[13] one of the non constructible vertex is characterized by:

$$\rho\left(\right) > 0$$


$$\rho\left(\right) > 0$$

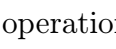

and all other configurations have zero probability. Thus, the set of all configurations of positive probability is

$$\overline{Z} = \left\{ \right.$$

$$\left. \right\}$$

while the set of configuration of zero probability is just the complement of $\overline{Z}$ in the set of all configurations on cluster  . We seek to cancel terms of the form $-\rho_\alpha(\gamma)$ with $\gamma \in \overline{Z}$ by terms of the form $\rho_\alpha(\gamma)$ with $\gamma \notin \overline{Z}$.

Suppose we want to cancel all negative terms generated by

$$-\rho\left(\right). \tag{5.31}$$

We first find the symmetry operation s such that cluster $s\left(\right)$  overlaps cluster  the most. Of course, if s is identity, the best overlap is obtained. But then it is impossible to find a configuration not in $\overline{Z}$ which fits on top of a configuration in $\overline{Z}$. We then try the next best thing: a s such that $s\left(\right) \cup$  is the cluster

 .

operation s such that we simultaneously have $\gamma_1 \cap (\alpha_1 \cap \alpha_2) = \gamma_2 \cap (\alpha_1 \cap \alpha_2)$ and $\gamma_1 \cap (\alpha_1 \cap s(\alpha_2)) = s(\gamma_2) \cap (\alpha_1 \cap s(\alpha_2))$ and $\alpha_1 \backslash \alpha_2 \neq \emptyset$ and $\alpha_1 \backslash s(\alpha_2) \neq \emptyset$ and $\gamma_1 \backslash \alpha_2 = \emptyset$ and $\gamma_1 \backslash s(\alpha_2) = \alpha_1 \backslash \alpha_2$. If this is not the case, one can show that either the probability of configuration γ_1 on cluster ∞ or of configuration $(\infty \backslash \alpha_1) \cup \gamma_1$ on cluster ∞ cannot be cancelled.

[13] We denote cluster sites which could host either $\circ$ or $\bullet$ atom by $\diamond$.

Let us expand Equation (5.31) on this larger cluster:

$$-\rho\left(\begin{smallmatrix}\circ\\ \bullet\circ\bullet\\ \circ\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\circ\\ \bullet\circ\bullet\\ \bullet\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\circ\\ \bullet\circ\bullet\\ \circ\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\circ\\ \bullet\circ\bullet\\ \bullet\end{smallmatrix}\right). \tag{5.32}$$

We can identify, on subcluster $\begin{smallmatrix}&&\diamond\\ \diamond\diamond\diamond\end{smallmatrix}$, a configuration which is not in $\overline{Z}$ for all terms but the first one. Therefore, by Proposition 8, we can cancel these terms by adding three terms to Equation (5.31):

$$-\rho\left(\begin{smallmatrix}\circ\\ \bullet\circ\bullet\end{smallmatrix}\right) + \left(\rho\left(\begin{smallmatrix}\\ \circ\bullet\circ\\ \bullet\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\\ \circ\bullet\bullet\\ \circ\end{smallmatrix}\right) + \rho\left(\begin{smallmatrix}\\ \circ\bullet\bullet\\ \bullet\end{smallmatrix}\right)\right). \tag{5.33}$$

One can check that, if we expand each of these terms on the larger cluster $\begin{smallmatrix}\diamond&&\diamond\\ &\diamond\diamond&\\ \diamond&&\diamond\end{smallmatrix}$, we obtain an expression which contains only one negative term: $-\rho\left(\begin{smallmatrix}\circ\\ \bullet\circ\bullet\\ \circ\end{smallmatrix}\right)$. This term cannot be cancelled using Proposition 8 because it has a configuration which is in $\overline{Z}$ on subcluster $\begin{smallmatrix}&&\diamond\\ \diamond\diamond\diamond\end{smallmatrix}$. This suggests a very simple scheme to directly find which negative terms are remaining without performing cancellations one at the time. All remaining negative terms are obtained by finding all possible ways of placing configurations belonging to $\overline{Z}$ on subcluster $\begin{smallmatrix}&&\diamond\\ \diamond\diamond\diamond\end{smallmatrix}$ such that the added atoms agree with the atoms already placed on subcluster $\begin{smallmatrix}\diamond\\ \diamond\diamond\end{smallmatrix}$. Hence, in our case, the only remaining negative term is $-\rho\left(\begin{smallmatrix}\circ\\ \bullet\circ\bullet\\ \circ\end{smallmatrix}\right)$.

Applying expansion summation and cancellation of identical terms lets us obtain another convenient result. We can rewrite Equation (5.33) in a shorter form:

$$-\rho\left(\begin{smallmatrix}\circ\\ \bullet\circ\bullet\end{smallmatrix}\right) + \left(\rho\left(\begin{smallmatrix}\\ \circ\bullet\end{smallmatrix}\right) - \rho\left(\begin{smallmatrix}\\ \circ\bullet\circ\\ \circ\end{smallmatrix}\right)\right). \tag{5.34}$$

It follows that, at any time, the state of the computations can be described by

- an inequality of the form of Equation (5.34) which is expressible in term of probability of occurrence of configurations on subclusters of the low cluster;
- a list of configurations on larger clusters corresponding to negative terms which have not yet been cancelled.

Let us now generalize these results. Start with a configuration in $\overline{Z}$ on a cluster $\alpha' = \alpha$. Find the symmetry s such that cluster $s(\alpha)$ overlaps α' the most. Add, to the inequality, the probability of occurrence of the configuration appearing on cluster $\alpha' \cap s(\alpha)$. Subtract, from the inequality, all the probability of occurrence of configurations in $\overline{Z}$ which we can place on $s(\alpha)$ without conflicting with the atoms already placed on α'. The negative terms we still have to cancel are obtained by overlapping the configuration on α' and all configurations in $\overline{Z}$ which we could place on $s(\alpha)$. Now set $\alpha' = \alpha \cup s(\alpha)$ and repeat the process until no more negative terms remain. More formally, we have the following.

Proposition 9 *Let $\overline{Z}$ be the set of configurations on some low cluster α which have a positive probability of occurrence. Suppose we want to remove as many negative terms as possible (in the sense of Equation (5.30)) in an expression of the form*

$$-\rho_{\alpha'}(\gamma')$$

by using only one symmetry operation s and the minimum number of positive terms. The optimum way to proceed is to write

$$-\rho_{\alpha'}(\gamma') + \rho_{\nu}(\gamma' \cap \nu) - \sum_{\gamma \in F} \rho_{s(\alpha)}(s(\gamma)) \tag{5.35}$$

where

- $\nu = \alpha' \cap s(\alpha)$
- $F = \left\{\gamma \in \overline{Z} \text{ such that } \gamma' \cap \nu = s(\gamma) \cap \nu\right\}$
- *s is*
  - *such that F is empty;*
  - *or, if this is not possible, such that $|F|$ is minimum under the constraint that $s(\alpha) \not\subset \alpha'$.*

Moreover, Equation (5.35) expands to

$$\sum_{\delta \in A} \rho_{\beta}(\delta) - \sum_{\delta \in B} \rho_{\beta}(\delta) \tag{5.36}$$

where

- $\beta = \alpha' \cup s(\alpha)$
- $A = \{\delta \subset \beta \text{ such that } (\delta \cap \alpha') \neq \gamma \text{ and } (s^{-1}(\delta) \cap \alpha) \notin F \text{ and } (\delta \cap \nu) = (\gamma \cap \nu)\}$
- $B = \{\delta \subset \beta \text{ such that } (\delta \cap \alpha') = \gamma' \text{ and } (s^{-1}(\delta) \cap \alpha) \in F\}$.

(Note that if F is empty, no negative term remains.)

5.6.8 Merging the Construction Process with Inequality Generation

The result of the previous section is very useful for two reasons. First, at a given vertex, there are generally only a few configurations γ of positive probability ($\gamma \in \overline{Z}$) even though there might be a large number of configurations with zero probability ($\gamma \in Z$). Hence, by expressing our method only in term elements of $\overline{Z}$, we have significantly reduced the complexity of our algorithm.

Second, we have merged the computation of new inequalities with the reconstruction process. To see this, we note that we obtain all negative terms by finding all

possible ways of overlapping configuration of positive probability so that the atom type match at the overlapping sites. This is exactly what one does to construct a structure from the probabilities. We even have the following result.

Proposition 10 *Repeated application of Proposition 9 is guaranteed to remove all negative term after a finite number of steps if and only if vertex v is non constructible.*
Proof: *If the process stops, we have found an inequality excluding the vertex v. If the process never stops, the negative terms we have not been able to cancel are probability of occurrence of configurations on clusters whose size tends to infinity.*[14] *These configurations are obtained by overlapping configurations of positive probability in all possible ways which makes identical atoms superpose. Hence, in the limit, each negative terms* $-\rho_\infty(\delta)$ *define a structure*[15] δ *represented by vertex v.* ■

In the proof of this proposition, we have assumed that once we have found a structure made only of configurations in $\overline{Z}$, this structure automatically has the correlations values obtained at vertex v. Since this is not obvious, we need to prove the following.

Proposition 11 *All structures associated with a vertex v of the low configurational polyhedron can be constructed by finding all possible ways to overlap low configurations having a positive probability of occurrence. (And the correlation values of the resulting structures will always match those of vertex v.)*
Proof: *Vertex v is uniquely defined by stating which low configurations have a zero probability (or equivalently, which low configurations have a positive probability). We attempt to construct vertex v by only using our knowledge of the set* $\overline{Z}$ *of all low configurations of positive probability. We need to explore the two possible alternatives.*

Suppose we have found a way to overlap configurations in $\overline{Z}$ *to build a structure and that all configurations in* $\overline{Z}$ *have a strictly positive probability of occurrence in the structure. Any constructible structure is associated to a point in the polyhedron. Yet, vertex v is the only point in the low polyhedron for which configurations in* $\overline{Z}$ *are the only one to have positive probability. Therefore, the structure obtained must necessarily have the appropriate correlations values.*

Suppose we have found a structure by overlapping configurations in $\overline{Z}$ *but that not all configurations in* $\overline{Z}$ *have a strictly positive probability of occurrence in the structure. Hence we have more configurations of zero probability than in the previous case. Any real structure must correspond to a point in the polytope. Yet, there are no points in the polytope which are defined by so many configurations of zero probability. Vertex v is the only point where all configurations not in* $\overline{Z}$ *have zero probability and we would need a point where even more configurations have zero probability. This is impossible.* ■

[14] *we can always make these clusters grow at comparable speed in every directions*
[15] *A structure is configuration defined on cluster* ∞

5.6.9 Example

Let us now apply Proposition 9 to prove non constructibility of one vertex in the case of the honeycomb "lattice". We recall that one of the non constructible vertex is defined by

$$\overline{Z} = \left\{ \begin{array}{c} \circ \\ \circ\bullet\circ \end{array} , \begin{array}{c} \circ \\ \bullet\circ\bullet \end{array} , \begin{array}{c} \bullet \\ \bullet\circ\circ \end{array} , \begin{array}{c} \bullet \\ \circ\circ\bullet \end{array} \right\}.$$

In the following example, we will show the current value of the inequality we are building at each application of Proposition 9. These inequalities will be expressed in term of configurations on subclusters of the low cluster (as in Equation (5.35)). We will not need the complete expression of these inequalities on larger clusters (as in Equation (5.36)). We will only keep track of the negative terms since these are the ones which pose a problem. In Figure 5-6, we list the configurations γ associated with negative terms $-\rho(\gamma)$ at each application of Proposition 9. This example shows only one of the many possible ways through which an inequality can be found. A practical algorithm would have to explore many possible alternatives.

In the following equations, "..." stands for all the terms of the previous inequality.

1. We start with some (arbitrarily chosen) configuration in $\overline{Z}$:

$$-\rho\left(\begin{array}{c} \circ \\ \bullet\circ\bullet \end{array} \right) \geq 0.$$

2. There many ways to overlap the next cluster. We choose among the ones which minimizes the number of negative terms (*i.e.*, which minimizes the number of alternative way to *decorate* the new cluster):

$$\ldots + \rho\left(\begin{array}{c} \circ \\ \bullet \end{array} \right) - \rho\left(\begin{array}{c} \circ \\ \circ\bullet\circ \end{array} \right) \geq 0.$$

3. We can do the same at the other similar locations:

$$\ldots + \rho\left(\begin{array}{c} \circ \\ \bullet \end{array} \right) - \rho\left(\begin{array}{c} \circ \\ \circ\bullet\circ \end{array} \right) \geq 0,$$

4.

$$\ldots + \rho\left(\begin{array}{c} \circ \\ \circ \end{array} \right) - \rho\left(\begin{array}{c} \circ \\ \bullet\circ\bullet \end{array} \right) \geq 0,$$

5.

$$\ldots + \rho\left(\begin{array}{c} \\ \bullet\circ \end{array} \right) - \rho\left(\begin{array}{c} \circ \\ \circ\bullet\circ \end{array} \right) \geq 0.$$

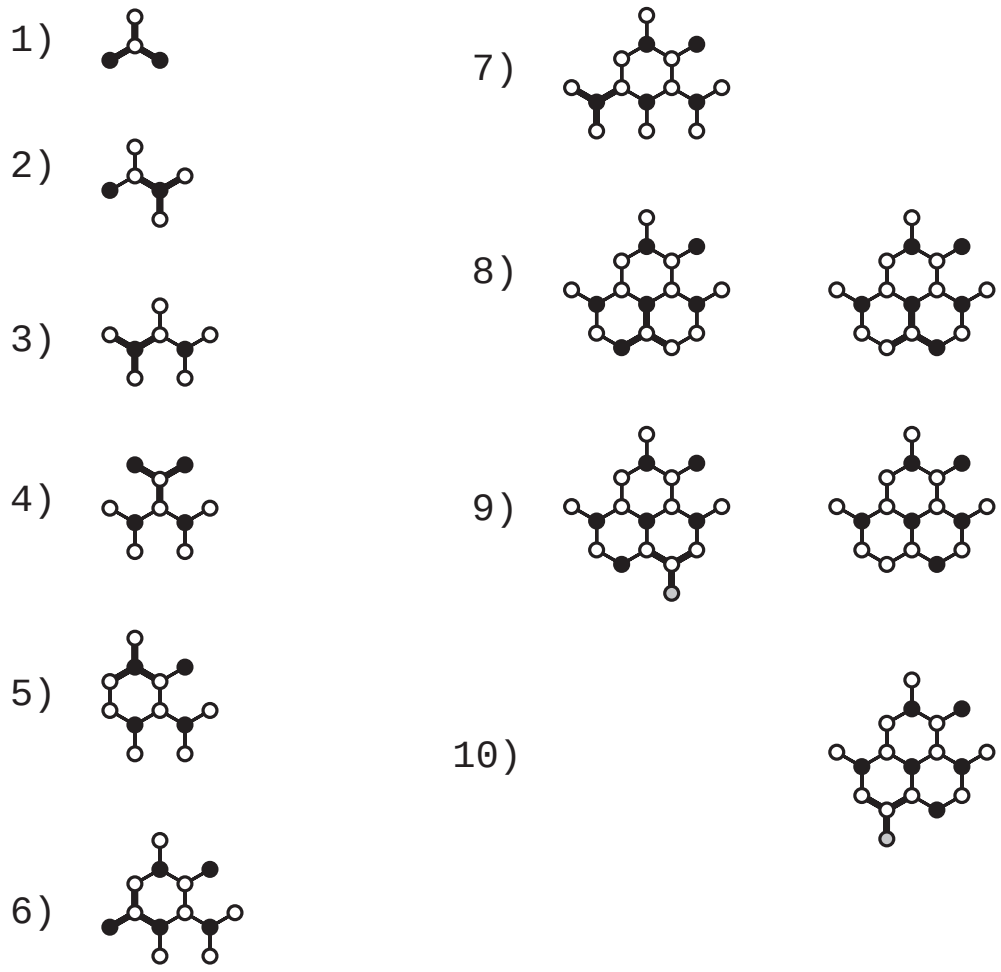


Figure 5-6: Example of the Use of Proposition 9 (see text).

6. We can now have a three sites overlap, which certainly gives the maximum amount of cancellation:

$$\ldots + \rho\left(\begin{array}{cc} \circ & \\ \circ & \bullet \end{array}\right) - \rho\left(\begin{array}{ccc} & \circ & \\ \bullet & \circ & \bullet \end{array}\right) \geq 0.$$

7. We are back to the usual two sites overlap:

$$\ldots + \rho\left(\begin{array}{cc} \circ & \\ \bullet & \end{array}\right) - \rho\left(\begin{array}{ccc} & \circ & \\ \circ & \bullet & \circ \end{array}\right) \geq 0.$$

8. There are now many different ways to place the next cluster which all seem as good as the other. We select the location which makes the growing structure as "round" as possible.[16] (We intuitively know that frustration is better detected by round cluster than by long and thin ones.)

$$\ldots + \rho\left(\begin{array}{cc} \bullet & \\ \circ & \end{array}\right) - \rho\left(\begin{array}{ccc} & \bullet & \\ \circ & \circ & \bullet \end{array}\right) - \rho\left(\begin{array}{ccc} & \bullet & \\ \bullet & \circ & \circ \end{array}\right) \geq 0.$$

Note that two configurations could fit in this case and we now have two negative terms to cancel.

9. The first negative term contains three adjacent $\circ$ atoms on one subcluster. No configuration belonging to $\overline{Z}$ could fit there. Hence we add the term $\rho\left(\begin{array}{ccc} & & \\ \circ & \circ & \circ \end{array}\right)$ without subtracting any other term. Moreover, we do not generate new negative terms (we encounter the case where "F is empty" in Proposition 9).

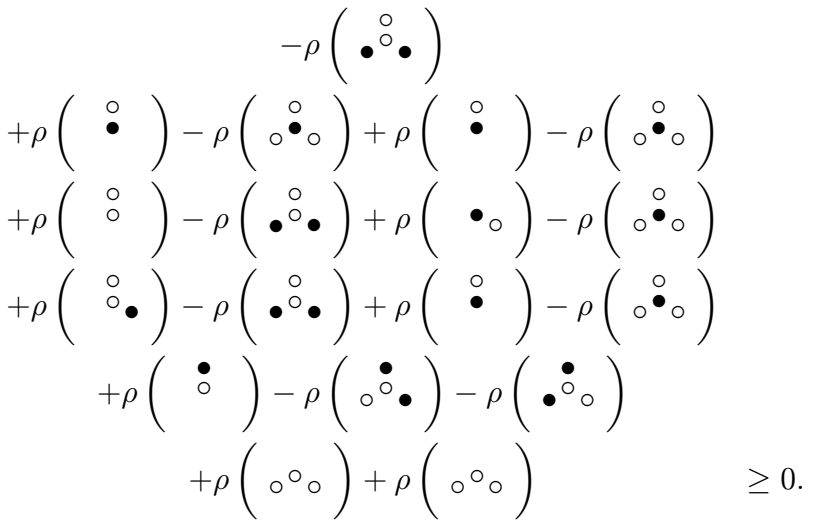
$$\ldots + \rho\left(\begin{array}{ccc} & & \\ \circ & \circ & \circ \end{array}\right) \geq 0.$$

10. The other negative term also has three adjacent $\circ$ atoms. For the same reason, we have:

$$\ldots + \rho\left(\begin{array}{ccc} & & \\ \circ & \circ & \circ \end{array}\right) \geq 0. \tag{5.37}$$

[16] "Roundness" can be evaluated by the ratio of the number of sites lying at the boundary of the growing structure to the total number of sites.

Therefore, the inequality



is valid and proves that vertex v is not constructible. We notice that what makes our procedure terminate agrees perfectly with our intuitive notion of frustration: finding a subcluster on which we are not able to fit any configuration belonging to $\overline{Z}$.

The procedure illustrated above is *guaranteed* to work in all cases. It either constructs a structure or constructs an inequality which proves non constructibility. Moreover, the complexity of the method is not worse than the one of the usual construction process from the probabilities. This is a lot faster than the conventional way of dealing with frustrations. In our example with the honeycomb "lattice", one would need a 13 point cluster to account for the frustration we have detected. The configurational polyhedron generated by this cluster has at least 1000 faces which makes the task of projecting this polyhedron to a low dimensional space intractable. In contrast, we were able to perform our computations by hand.

5.6.10 Improved Method

The method presented in the last section is not perfect. It can generate inequalities which prove non constructibility but it is not guaranteed to generate all the tightest inequalities one would need to find the actual ground state.

The source of the weakness is that we have neglected all the positive terms generated when we apply Proposition 9. These positive terms could be used to cancel some of the negative terms without adding *new* positive terms. The problem is that there are *many* positive terms. Fortunately there is a short hand way to represent them.

Let us take the example which lead to Equation (5.34):

Can we find a simple expression for the positive terms generated by this expression when we expand each term to cluster  ? Yes, we simply note the following.

- All positive terms come from the second term and they must be of the form $\rho\left(\begin{smallmatrix}\diamond & & \diamond\\ & \circ\bullet & \\ \diamond & & \diamond\end{smallmatrix}\right)$ where $\diamond$ represents either $\circ$ or $\bullet$.
- They cannot be of the form $\rho\left(\begin{smallmatrix}\circ & & \diamond\\ & \circ\bullet & \\ \bullet & & \diamond\end{smallmatrix}\right)$ because they would be cancelled by the first term.
- They cannot be of the form $\rho\left(\begin{smallmatrix}\diamond & & \circ\\ & \circ\bullet & \\ \diamond & & \circ\end{smallmatrix}\right)$ either because they would be cancelled by the last term.

These three facts exactly define all positive terms and require less memory to store than the 9 positive terms themselves. In fact, in Proposition 9, the set A is defined in this way. Any positive term generated by Proposition 9 can be expressed in this form: two ends which should not fit a given configurations and a middle part which should fit a given configuration.

Another attractive property of this representation is that it is easy to detect if a positive term expressed in this form cancels negative terms. We will not delve into the details of this procedure because it is similar in concept to what we have done so far, except that there are more cases to handle.

Taking the positive terms into account makes the algorithm somewhat more complex. Before adding new positive term, we scan through existing positive terms to see if they would fit somewhere. It is just as if, beside set $\overline{Z}$, we had access to other configurations to fit in our growing structure.

The order of complexity of the algorithm does not suffer too much from this addition. For example, in our example of the honeycomb "lattice", it is still possible to perform the computation by hand to generate inequalities which enable us to find 2 of Kanamori's ground states [41]. The ground states we found would have required a 13 point cluster by traditional means.

To make sure this method can be applied to more complex problems we need to handle the so-called *branching problem*.

5.6.11 Branching

As we attempt to construct a structure, we have to consider a lot of different alternatives. We can branch on

- where to overlap the next cluster;
- which configuration to use to decorate that new cluster;
- which configuration to subtract (at the initial step, and maybe later on, as was explained in Section 5.4.4).

The second type of branching is essential. We really have to explore all cases: every possibility corresponds to a negative term which *has* to be cancelled. However, the two other types of branching are facultative: if we had a way to predict which branch will yield the tightest inequality, we could completely avoid branching.

We can use our measure M of how close we are to the complete removal of all negative terms (see Section 5.6.5) as a guide to select the most promising branches. Of course, this scheme makes the method approximative: some apparently non promising alternatives may turn out to be promising later on. We thus keep a list of the p most promising alternatives instead of keeping only *the* most promising alternative. At each step, we perform a small amount of computation in every promising branch. When a new facultative branching occurs, the least promising branches are deleted to make room for the new, more promising branches. We never explore more than p alternatives at any time. The situation is analogous to having p identical programs running in parallel, each one testing a different branching sequence. As we had discussed in Section 5.4.4, the first program to yield a valid inequality will yield a tighter inequality than the other programs when they eventually find one. We just hope that an exact answer is still possible without using a too large p.

5.7 Conclusion

We introduced a method which is capable of proving either the constructibility or the non constructibility of a given vertex. Its order of complexity is comparable with the one of the usual construction process.

A more involved algorithm seems to be able to find all valid inequalities which exclude a given vertex. It is guaranteed that our method will never fail by giving us a wrong inequality. The only possible problem could be that it may not be able to give us all the inequalities we need. From a theoretical point of view, we are missing a rigorous proof that all inequalities which can be obtained by a cluster of finite size are obtainable by our method. The only reason why this could be not true was the subject of Section 5.6.6. But the circumstances leading to a failure of our method seem very unlikely. (No counterexample has been found yet.)

Although our method appears to save a lot of computations, it still has an exponential order of complexity. However, we have a way to control the computation time by only exploring the most promising alternatives in a tree of possible computations leading to valid inequalities. The challenge is to devise a program that possesses the required "intuition" to explore only these promising options.

For at least one example, the capabilities of this new method are surprising. We were able to obtain a 2 of Kanamori's ground state on the honeycomb "lattice" (see, for instance, [41]). These computations were performed manually where as the computation of the v-matrix for a cluster large enough to account for the same frustration effects took about 20 minutes of computer time.

Chapter 6

Conclusion

A long standing limitation of first-principles phase diagram calculations is the difficulty to account for the impact of lattice vibrations on phase stability. A thorough review of the literature on this topic has been presented.

In this thesis, a clear picture of the origin of vibrational entropy differences between phases in an alloy system has been developed. Vibrational entropy change can be attributed to the changes in bond stiffness associated with the changes in bond length that take place during a phase transformation. This understanding was made possible by accurate first-principles calculations of vibrational entropy differences. These calculations enabled us to develop controlled approximations that allow the inclusion of vibrational effects in phase diagram calculations at a reasonable computational cost.

This thesis also address an additional issue regarding the calculation of vibrational properties from first principles. The use of the well-known Local Density Approximation results in a systematic underprediction of lattice parameters. We show that this bias, which has an important impact on calculated vibrational properties, can be corrected through a simple semi-empirical scheme. The success of such a simple approach also brings useful insight into the behavior of the exact exchange-correlation potential in a solid state system, a quantity that is central to Density Functional Theory.

This thesis brought light onto another intricate problem associated with phase diagram calculations: the ground state problem. Although the technique to find ground state is well established, its practical implementation remains computationally demanding. A new useful connection between the construction of the so-called "configurational polytope" and the geometric problem of constructing a structure that has a given correlation vector has been presented. There is good hope that such intuition will lead to practical solution to the ground state problem.

While very useful results have been obtained in all the aspects explored, the main contribution of this thesis is clearly the development of the "bond stiffness vs. bond length" interpretation which both summarizes the key phenomenon driving vibrational entropy changes and provides a practical tool to model them.

Appendix A

Selected topics related to lattice vibrations in alloys

A.1 The absence of mass effects in the high-temperature limit

In the high-temperature limit, the vibrational entropy is determined by the product of the frequencies of all normal modes of vibrations ν_m, which can be related to the eigenvalues λ_m of the $3N \times 3N$ dynamical matrix D of the system (up to a constant):[1]

$$\begin{aligned}\sum_m \ln(\nu_m) &= \ln\left(\prod_m \nu_m\right) = \frac{1}{2}\ln\left(\prod_m \lambda_m\right) + \text{const} \\ &= \frac{1}{2N}\ln(\det D) + \text{const}\end{aligned}$$

Using the properties of determinants, we can write:

$$\begin{aligned}\frac{1}{2N}\ln(\det D) &= \frac{1}{2N}\ln\left(\det\left(M^{1/2}\Phi M^{1/2}\right)\right) \\ &= \frac{1}{2N}\ln(\det(\Phi)\det(M)) = \frac{1}{2N}\ln\left(\det(\Phi)\prod_i M_i^3\right) \\ &= \frac{1}{2N}\ln(\det(\Phi)) + \frac{3}{2N}\sum_i \ln M_i\end{aligned}$$

where M is the $3N \times 3N$ diagonal matrix of all the N atomic masses of the system (each repeated three times) while Φ is $3N \times 3N$ the matrix obtained by regrouping all the 3×3 force constant tensors $\Phi(i,j)$ in a single matrix (analogously to Equation

[1] To simplify the exposition and avoid the problem that the dynamical matrix has three zero eigenvalues associated with the possibility of a rigid translation of the system, we assume that some of the atoms of the system are attached to a fixed point of reference by a weak spring. In the thermodynamic limit, this assumption becomes inconsequential.

(2.11)). Now consider the change in the value of $\sum_m \ln(\nu_m)$ when an N_A atoms of type A and N_B atoms of type B are combined to form an alloy. Let the subscripts AB, A and B respectively denote the properties of an $A_{(N_A/N)}B_{(N_B/N)}$ alloy, a pure crystal of element A and a pure crystal of element B.

$$\begin{aligned}
\Delta\left(\sum_m \ln(\nu_m)\right) &= \sum_m \ln\left(\nu_m^{AB}\right) - x_A \sum_m \ln\left(\nu_m^{A}\right) - x_B \sum_m \ln\left(\nu_m^{B}\right) \\
&= \frac{1}{2}\ln\left(\det\left(\Phi^{AB}\right)\right) + \frac{3}{2}N_A \ln M_A + \frac{3}{2}N_B \ln M_B \\
&\quad -\frac{1}{2}\ln\left(\det\left(\Phi^{A}\right)\right) - \frac{3}{2}N_A \ln M_A - \frac{1}{2}\ln\left(\det\left(\Phi^{B}\right)\right) - \frac{3}{2}N_B \ln M_B \\
&= .\frac{1}{2}\ln\left(\det\left(\Phi^{AB}\right)\right) - \frac{1}{2}\ln\left(\det\left(\Phi^{A}\right)\right) - \frac{1}{2}\ln\left(\det\left(\Phi^{B}\right)\right)
\end{aligned}$$

All the terms involving masses exactly cancel one another.

A.2 A simple model of anharmonicity

Two assumptions are made. First, the elastic energy of the motionless lattice is assumed quadratic in volume:

$$E^*(V) = \frac{B}{2V_0}(\Delta V)^2$$

where B is the bulk modulus, V_0 the equilibrium volume at $0K$ (ignoring zero-point motion) and $\Delta V = V - V_0$. Second, the high temperature limit of the vibrational free energy is used:

$$F_{vib}(T,V) = k_B T \sum_i \ln\left(\frac{h\nu_i}{k_B T}\right)$$

In this approximation, the volume-dependence of F_{vib} takes on a particularly simple form:

$$\frac{\partial F_{vib}(T,V)}{\partial V} = \frac{3Nk_B T\overline{\gamma}}{V}$$

where

$$\overline{\gamma} = \frac{1}{3N}\sum_{i=1}^{3N}\frac{V}{\nu_i}\frac{\partial \nu_i}{\partial V}$$

is an average Grüneisen parameter. In the high-temperature limit, an average Grüneisen parameter can easily be defined, because the population of the phonon modes is no longer temperature-dependent, and any change in entropy can be unambiguously attributed to shifts in phonon frequencies. At lower temperatures, the changes in

phonon population would need to be accounted for as well.

If we assume that the volume-dependence of the vibrational free energy is linear in volume, we have:

$$\begin{aligned} F(T,V) &= E^*(V) + F_{vib}(T,V) \\ &= E^*(V) + F_{vib}(T,V_0) + \left.\frac{\partial F_{vib}}{\partial V}\right|_{V=V_0} \Delta V \\ &= \frac{B}{2V_0}(\Delta V)^2 + F_{vib}(T,V_0) + \frac{3Nk_BT\overline{\gamma}}{V_0}\Delta V. \end{aligned}$$

Minimizing this expression with respect to ΔV yields:

$$\frac{\Delta V}{N} = \frac{3k_BT\overline{\gamma}}{B}$$

where $\frac{3k_BT\gamma}{B}$ is the coefficient of volumetric thermal expansion. The resulting temperature dependence of the free energy (for one given configuration $\vec{\sigma}$) is given by

$$\frac{F(T)}{N} = \frac{F(T,V_0)}{N} - \frac{(3k_BT\overline{\gamma})^2}{2B(V_0/N)}.$$

It is interesting to note that half of the vibrational free energy decrease due to thermal expansion is cancelled by the energy increase of the motionless lattice. Hence, vibrational entropy differences originating from differences in thermal expansion between phases have, relative to other sources of vibrational entropy changes, half the effect on phase stability.

A.3 Modeling the disordered state

Although in phase diagram calculations, the use of the cluster expansion bypasses the problem of directly calculating the vibrational entropy of a disordered phase, there are cases where it is of interest to directly calculate the vibrational properties of the disordered state. For instance, in studies that seek to assess the importance of lattice vibrations [8, 115, 136, 106], it is instructive to compute the vibrational entropy change upon disordering an alloy, since this quantity can be straightforwardly used to estimate the effect of lattice vibrations on transition temperatures with the help of Equation (2.4). Here are the most common methods used tomodel the disordered state.

Perhaps the most obvious and brute force approach to modeling the disordered state is the use of a large supercell where the occupation of each site is chosen at random. This approach was chosen in all EAM calculations [2, 8, 115, 90] as well as in other investigations [121]. Unfortunately, it is generally not feasible in the case of *ab initio* calculations.

The virtual crystal approximation (VCA) consists of replacing each atom in a disordered alloy by an "average" atom whose properties are determined by a con-

centration weighted average of the properties of the constituents. In the limit where the chemical species have nearly identical properties, this approximation is justified. This model has been commonly used to interpret neutron scattering measurements of phonon dispersion curves in the case of disordered alloys [49, 95]. It has also been used in a some theoretical investigations [32, 110]. However, the virtual crystal approximation has been repeatedly shown to be insufficiently accurate for the purpose of calculating differences in vibrational entropies between distinct compounds [8, 115, 121, 96]. Its weaknesses are numerous: It is unable to model "bond proportion" effects, volume effects and local relaxations. It also fails to give a mass-independent high temperature limit.

Special Quasirandom Structures (SQS) [151] combine the idea of cluster expansion with the use of supercells. SQS are the periodic structures that best approximate the disordered state in a unit cell of a given size. The quality of a SQS is described by the number of its correlations that match the ones of the true disordered state. There is thus a direct connection between cluster expansions and SQS: If there exists a truncated cluster expansion that is able to predict properties.of the disordered state there exists an SQS that provides an equally accurate description of the disordered state.

SQS have been very successfully used to obtain electronic and thermodynamic properties of disordered materials (see, for example, [63]).The accuracy of the SQS approach in the context of phonon calculations has been benchmarked using embedded atoms potentials which allow for the comparison with the "exact" vibrational entropy of the disordered state with a large supercell [90]. It has been found that, for the purpose of calculating vibrational properties, an SQS having only 8 atoms in its unit cell already provide a good approximation of the disordered state in the case of an fcc alloy at concentration 3/4. While the performance of this small SQS is remarkable in a model system where local relaxations are disallowed, it tends to degrade somewhat when relaxations are allowed to take place. This effect can naturally be explained by the fact that relaxations are known to introduce important multibody terms in the cluster expansion, which translates into the requirement that the SQS must correctly reproduce the corresponding multibody correlations.

The success of small SQS opened the way for the use of more accurate energy models to calculate vibrational properties of disordered alloys. SQS have been applied to the *ab-initio* calculation of vibrational entropy in disordered Ni_3Al and Pd_3V alloys [136, 137]

A.4 The Einstein model

In the Einstein model of a solid, the free energy, in the high temperature limit, is given by

$$\begin{aligned} F &= k_B T \ln\left(\frac{h}{k_B T}\prod_i \nu_i\right) \\ &= \frac{k_B T}{2}\ln\left(\frac{h}{k_B T}\det D\right) \\ &= \frac{k_B T}{2}\ln\left(\frac{h}{k_B T}\det\left(M^{1/2}\Phi M^{1/2}\right)\right) \\ &= \frac{k_B T}{2}\ln\left(\frac{h}{k_B T}\det M\det\Phi\right) \end{aligned}$$

where D and Φ are, respectively, the $3N\times 3N$ dynamical matrix and force constant matrix of the system while M is the matrix of the masses:

$$M_{ij} = \delta_{ij} M_j.$$

It can be shown [134] that for any positive definite matrix Φ

$$\det\Phi \le \prod_i \Phi_{ii},$$

implying that

$$F \le \frac{k_B T}{2}\ln\left(\frac{h}{k_B T}\prod_i M_i\Phi_{ii}\right)$$

where the right-hand side expression is nothing but the free energy of the system in the Einstein approximation. A lower bound can be obtained by a similar technique, by using the inverse of the force constant matrix

$$\det\Phi \ge \left(\prod_i \left(\Phi^{-1}\right)_{ii}\right)^{-1}.$$

The interpretation of the inverse Φ is simple: It is the matrix that maps forces F exerted on the atoms to the resulting displacements u of the atoms.

$$u = \Phi^{-1}F$$

While Φ_{ii} is related to the oscillation frequency of a single atom when all other atoms are held in place, $\left(\left(\Phi^{-1}\right)_{ii}\right)^{-1}$ is related to the oscillation frequency of an atom i when all surrounding atoms are allowed to relax so that the force exerted on them remains zero as atom i moves. Atom i has mass M_i while all other atoms are considered

massless and relax instantaneously. Atoms located infinitely far away from atom i are held in place with an infinitesimal force.

In conclusion, the free energy of a system is bounded by above and by below by the free energy of two Einstein-like systems:

$$\frac{k_B T}{2} \ln \left(\frac{h}{k_B T} \prod_i M_i \left(\left(\Phi^{-1} \right)_{ii} \right)^{-1} \right) \leq F \leq \frac{k_B T}{2} \ln \left(\frac{h}{k_B T} \prod_i M_i \Phi_{ii} \right)$$

A.5 Derivation of the "bond proportion" model

This appendix shows that, in an important class of systems, the bond proportion model is in fact the first order approximation to the true change in vibrational entropy induced by a change in the proportion of the different types of chemical bonds.

The alloy system is assumed to satisfy the following conditions:

- the high temperature limit is appropriate;
- the nearest-neighbor force constants can be written as $\Phi(i,j) = k_{\sigma_i \sigma_j} \phi(i,j)$ where $k_{\sigma_i \sigma_j}$ denotes the (scalar) stiffness of the spring connecting sites i and j with occupations σ_i and σ_j while the $\phi(i,j)$ are dimensionless spring constant tensors. The $\phi(i,j)$ are assumed equivalent under a symmetry operation of the space group of the parent lattice;
- all force constants $k_{\sigma_i \sigma_j}$ are such that

$$\frac{k_{\sigma_i \sigma_j}}{\sqrt{k_{\sigma_i \sigma_i} k_{\sigma_j \sigma_j}}} \ll 1.$$

Consider a d-dimensional solid made of N atoms connected by springs of characterized by symmetrically equivalent tensors $k\phi(i,j)$. Without loss of generality, the masses of all atoms are set to unity since the formation entropies in the high temperature limit are independent of the atomic masses (see Appendix A.1). In the high temperature limit, the vibrational free energy per atom is given by:

$$F_{\text{vib}} = \frac{k_B T}{2N} \sum_m \ln \lambda_m \tag{A.1}$$

where the sum is taken over the nonzero eigenvalues λ_m of the dynamical matrix D of the system. (The zero eigenvalues correspond the modes where the whole crystal moves rigidly. In the thermodynamic limit, these few missing degrees of freedom are inconsequential.)

Because all springs in the system are equivalent to each other, matrix D can be written as

$$D = kC$$

where C is a matrix of dimensionless geometrical factors independent of k but specific to the type of lattice. From this expression of D, it follows naturally that eigenvectors of D are independent of k and that its eigenvalues can be written as

$$\lambda_m = k l_m$$

where the l_m are geometric factors independent of k.

Consider what happens to S_{vib} when the stiffness of one of the springs is changed from k to $k + \Delta k$. Let ΔD denote the corresponding change in matrix D. To the first order, the resulting changes in the eigenvalues are given by:

$$\Delta\lambda_m = u'_m \Delta D u_m$$

where u_m is the (dimensionless) eigenvector of D associated with eigenvalue λ_m. Since D is linear in the spring constants, we can write

$$\Delta D = \Delta k \; B$$

where B is matrix of geometrical factors independent of k and Δk but specific to the type of lattice. While matrix B also depends on which spring is being modified, the matrices B corresponding to each spring are equivalent under a symmetry operation of the crystal's space group. The changes in the eigenvalues can then be expressed as:

$$\begin{aligned} \Delta\lambda_i &= \Delta k u'_m B u_m \\ &\equiv \Delta k \; g_m \end{aligned}$$

where g_i is a dimensionless number independent of k and Δk.

Substituting these results into Equation (A.1), we obtain:

$$F_{\text{vib}} = \frac{k_B T}{2N} \sum_m \ln\left(k l_m + \Delta k \; g_m\right).$$

To the first order, we can express the vibrational entropy change as

$$\begin{aligned} \Delta F_{\text{vib}} &= \left.\frac{\partial F_{\text{vib}}}{\partial \Delta k}\right|_{\Delta k = 0} \Delta k \\ &= \frac{k_B T}{2N} \sum_m \frac{g_m}{k l_m} \Delta k \\ &= k_B T \left(\frac{1}{2N} \sum_m \frac{g_m}{l_m}\right) \frac{\Delta k}{k} \\ &\equiv k_B T G \frac{\Delta k}{k}. \end{aligned}$$

where G is a dimensionless geometrical factor depending only on the lattice type.

In the limit of $\Delta k \ll k$, we can obtain the change in vibrational entropy due to a change in all the spring constants by simply summing the effect of the change Δk_s in the stiffness of each spring s:

$$\Delta F_{\text{vib}} = k_B T G \sum_s \frac{\Delta k_s}{k} \tag{A.2}$$

To determine the value of G, we consider the following particular case for which the exact vibrational entropy change is known. Once the value of G is known, it can be used in any other case sharing a same lattice type.

In a solid bound by springs of stiffness k is given by, if the stiffness of all springs is increased by Δk, each eigenvalue λ_m becomes $\lambda_m \frac{k+\Delta k}{k}$ and the vibrational entropy becomes:

$$\begin{aligned}
F'_{\text{vib}} &= \frac{k_B T}{2N} \sum_i \ln\left(\lambda_i \frac{k+\Delta k}{k}\right) \\
&= \frac{k_B T}{2N} \sum_i \left(\ln \lambda_i + \ln \frac{k+\Delta k}{k}\right) \\
&= F_{\text{vib}} + \frac{k_B T}{2N} \sum_i \ln \frac{k+\Delta k}{k} \\
&\approx F_{\text{vib}} + k_B T \frac{Nd}{2N} \frac{\Delta k}{k} + O\left((\Delta k)^2\right) \\
&= F_{\text{vib}} + k_B T \frac{Nd}{2N} \frac{1}{ZN/2} \sum_s \frac{\Delta k}{k} \\
&= F_{\text{vib}} + \frac{k_B T d}{ZN} \sum_s \frac{\Delta k}{k}.
\end{aligned}$$

where Z is the number of nearest neighbors and $\sum_s$ denotes a sum over all nearest neighbor bonds. Since this result is exact to the first order, we can compare it to Equation (A.2) and identify the unknown constant G to be $\frac{d}{ZN}$. We thus obtain the following result:

$$\Delta F_{\text{vib}} = \frac{3 k_B T}{ZN} \sum_s \frac{\Delta k_s}{k}. \tag{A.3}$$

We now turn to the problem of calculating the vibrational entropy of mixing in a binary alloy. We first define a normalized $3N \times 3N$ dynamical matrix $\hat{D}$ as follows:

$$\hat{D}_{\alpha\beta}(i,j) = \frac{\Phi_{\alpha\beta}(i,j)}{\sqrt{k_{\sigma_i\sigma_i} k_{\sigma_j\sigma_j}}}$$

where $k_{\sigma_i\sigma_i}$ is the spring constant of an $A-A$ bond if site i is occupied by a A atom similarly for a site occupied by a B atom. For the purpose of calculating free energy of formation, this normalized dynamical matrix gives the same result as the usual

dynamical matrix because the factors in the denominator exactly cancel out, for the same reason masses cancel out (See Appendix A.1). This transformation normalizes the spring constant associated with $A-A$ bonds and $B-B$ bonds to 1 while the spring constant associated with $A-B$ bond becomes $\left(k_{AB_j}/\sqrt{k_{AA}k_{BB}}\right)$ where k_{AB}, k_{AA} and k_{BB} respectively denote the true spring constants of $A-B$, $A-A$ and $B-B$ bonds. The usefulness of this normalization is to extend the applicability of Equation (A.3) to the case where k_{AA} and k_{BB} are very different.

Let us start with a phase separated mixture of A and B atoms. Let us think of this system as one where all atoms are identical but where the springs connecting them can be either one of three types $A-A$, $B-B$ or $A-B$. Where the springs are placed defines which type of atom sits at each site. We now replace one $A-A$ bond in the pure A phase by an $A-B$ bond and one $B-B$ bond in the pure B phase by an $A-B$ bond. By Equation (A.3), the resulting change in vibrational entropy per atom is:

$$\Delta F_{\text{vib}} = \frac{k_B T d}{ZN}\left(\frac{k_{AB}}{\sqrt{k_{AA}k_{BB}}} - 1 + \frac{k_{AB}}{\sqrt{k_{AA}k_{BB}}} - 1\right).$$

To satisfy the assumptions of the above derivation, we require that $k_{AB}/\sqrt{k_{AA}k_{BB}} \ll 1$. If we create a total number n_{AB} of $A-B$ bonds, we perform the above operation $n_{AB}/2$ times and the vibrational entropy change is:

$$\Delta F_{\text{vib}} = \frac{n_{AB}}{2}\frac{k_B T d}{ZN}\left(\frac{k_{AB}}{\sqrt{k_{AA}k_{BB}}} - 1 + \frac{k_{AB}}{\sqrt{k_{AA}k_{BB}}} - 1\right)$$

To the first order (when $k_{AB}/\sqrt{k_{AA}k_{BB}} \ll 1$), this expression is equivalent to

$$\Delta F_{\text{vib}} = \frac{n_{AB}}{N}\frac{k_B T d}{2Z}\ln\left(\frac{k_{AB}^2}{k_{AA}k_{BB}}\right).$$

The nearest neighbor ECI of the cluster expansion of the vibrational free energy is thus:

$$V_{1nn} = \frac{d}{8}k_B T \ln\left(\frac{k_{AA}k_{BB}}{k_{AB}^2}\right).$$

A.6 Instability

An extreme case of anharmonicity occurs when the energy surface, in the neighborhood of a configuration $\vec{\sigma}$, has no local minimum. As noted in [37] and [36], this situation occurs sufficiently frequently to deserve a particular attention. A typical example of such a situation occurs when the fcc-based $L1_0$ structure is unstable with respect to a deformation along the Bain path, which leads to a bcc-based $B2$ structure. While it is possible to construct a separate cluster expansion for the fcc and bcc phases, the fundamental question that arises is: What is the free energy of the $L1_0$ structure? Since it is unstable, the standard harmonic expression for the free energy

can clearly not be used.

One suggested solution to this problem, described in [36], is to perform the coarse graining in a different order than presented in Section 2.2.3. The sum over configurations is performed first, and the vibrational properties of the configurational averaged alloy are then calculated. The main limitation of this approach is that it would be extremely difficult to compute the averaged vibrational properties by any other method than by the so-called virtual crystal approximation (see Section A.3). Another limitation is that it only addresses instabilities with respect to cell shape distortions, ignoring instabilities with respect to internal degrees of freedom (*i.e.* atomic positions).

In this section, we present another approach to solve the instability problem. We argue that the general formalism developed in Section 2.2.3 can in fact be adapted to allow for instability.

While the coarse graining technique is most naturally interpreted as integrating out the "fast" degrees of freedom (*e.g.* vibrations) before considering "slower" ones (*e.g.* configurational changes)[28], the time scale of the various types of excitations is, in fact, irrelevant. The partition function is simply a sum over states which can be calculated in any order. As long as we can associate any vibrational state v of the system with a configuration $\vec{\sigma}$, the coarse graining procedure remains valid.

Under this point of view, it is clear that it does not matter whether there is even a local minimum of energy in the portion of phase space associated with configuration $\vec{\sigma}$. What is important, however, is that the neighborhood of configuration $\vec{\sigma}$ in phase space is thoroughly sampled (*i.e.* that the constrained system is ergodic) over a macroscopic time scale. There is no need for ergodicity within the time scale of the configurational excitations. If the neighborhood of a given configuration $\vec{\sigma}$ is not fully sampled before the alloy jumps to another configuration $\vec{\sigma}'$, it is still possible that the unsampled portion of phase space around $\vec{\sigma}$ will be visited at a later time, when the system returns in the neighborhood of configuration $\vec{\sigma}$. The ergodicity requirement at the macroscopic time scale imposes the important but intuitively obvious constraint that the phase space neighborhood of configuration $\vec{\sigma}$ cannot contain states that are associated to different phases of the system.

This discussion shows that there is no fundamental limitation to the applicability of the standard coarse-graining framework in the presence of instability. However, we still need to describe how the free energy of an unstable configuration could be determined in practice. The task is simplified by the fact that the free energy of an unstable stable does not need to be extremely accurately determined, because unstable states are relatively rarely visited, even at high temperatures. Nevertheless, it is important to assign a free energy to those unstable states, to ensure that the Ising model used to represent the alloy is well-defined.

The free energy associated with one configuration can be obtained by integrating $\exp\left[-\beta E(\vec{\sigma}, v)\right]$ with respect to v over the portion of phase space associated with $\vec{\sigma}$. In the classical limit, we can label the vibrational states v by the position each particle takes and the integration limits can be found by geometrical arguments. The quantum

mechanical equivalent of this operation is complex,[2] but unlikely to be needed in practice. The unstable states are essentially never visited at low temperatures, where a quantum mechanical treatment would be essential [3].

Focusing on the classical limit, we consider a unstable configuration $\vec{\sigma}$. Let D be the dynamical matrix evaluated at the saddle point of the energy surface closest to the ideal undistorted configuration $\vec{\sigma}$.[4] We consider that when the state v of the system is such that one atom moves away from its position at the saddle point by more than r, it should not longer be considered part of configuration $\vec{\sigma}$. For an instability with respect to internal degrees of freedom (atomic positions), a natural choice for r would be half the average nearest neighbor interatomic distance. For an instability with respect to unit cell deformation, r could be half the change in the average nearest neighbor distance induced by the displacive transformation.

The boundedness of the portion of phase space associated with $\vec{\sigma}$ can be accounted for by replacing the usual classical partition function associated with one normal mode of oscillation i by

$$\frac{1}{h}\int_{-L_i}^{L_i}\exp\left(-\frac{1}{2}\beta\,\dot{s}^2\right)d\dot{s}\int_{-L_i}^{L_i}\exp\left(-\frac{1}{2}\beta\lambda_i s^2\right)ds$$

where λ_i is the i-th eigenvalue of the dynamical matrix, h is Planck's constant and L_i is a measure of the size of the phase space neighborhood of $\vec{\sigma}$ along the direction associated with normal mode i. This size parameter can be expressed in terms of the parameter r just introduced. Let $u_i(j) = \frac{e_i(j)}{\sqrt{M_j}}$ where e_i is the i-th eigenvector of D and M_j is the mass of atom j. After normalizing u_i so that $\sum_j u_i^2(j) = N$, the number of atom in the system, we can then write

$$L_i = r\left(\max_{\text{nn } j,j'}\|u_i(j) - u_i(j')\|\right)^{-1}$$

where the maximum is taken over all nearest neighbor pairs of atoms j, j'. This choice of integration bounds approximately defines a neighborhood of $\vec{\sigma}$ such that no atom moves farther than r from its position at the saddle point (relative to its neighbors).

[2]The quantum partition function can be written as the trace of the matrix $\exp(-\beta H)$, where H is the (multibody) Hamiltonian of the system. The trace can computed in any convenient basis and in particular one could use Dirac delta functions. In this fashion, it is possible to define a localized free energy by summing only over the delta functions located in the neighborhood of one configuration $\vec{\sigma}$.

[3]This observation is related to the fact that quasi-harmonic approximation, which allows a quantum-mechanical treatment, is accurate up to a temperature where the classical limit is reached.

[4]Is is possible that an unstable configuration $\vec{\sigma}$ cannot be associated with a saddle point and the derivation would have to be modified. In particular the bounds of integration would have to be made asymmetric.

In this approximation, the free energy of an unstable state is given by:

$$\frac{F}{N} = \frac{E^*}{N} - \frac{k_B T}{N} \sum_i \ln \left(\frac{k_B T}{h\nu_i} \mathrm{erf} \left(L_i \sqrt{\frac{2 \left(\pi \nu_i \right)^2}{k_B T}} \right) \right)$$

where ν_i is the frequency of normal mode i and where the error function for real or imaginary arguments is given by

$$\mathrm{erf}\left(u\right) = \frac{2u}{\sqrt{\pi}} \int_0^1 e^{-u^2 s^2} ds.$$

The suggested definition of the free energy of an unstable configuration has interesting properties. First, as the neighborhood size L_i increases, the expression reduces to the usual harmonic expression. The effect of the correction is not limited to unstable modes: Modes that are so soft that it is likely that the motion of the atoms exceeds r are also affected. There may obviously be other definitions of L_i. The above example simply gives an example of how it could be calculated.

Going back to our initial example of the $L1_0 \rightarrow B2$ instability, we can now outline how this problem could be handled within the traditional coarse graining scheme. Two separate clusters expansion need to be constructed, one for the bcc phases and one for the fcc phases. But since we now know how to assign a free energy to the unstable $L1_0$ configuration, the fcc cluster expansion can be successfully defined. The free energy attributed to the $L1_0$ configuration should be sufficiently high so that the free energy curve of the fcc phase in the vicinity of 0.5 concentration will lie above the free energy curve of the bcc phase, as it should. The fact that both CVM or Monte Carlo calculations on the fcc lattice would attribute a positive probability to $L1_0$-like structures should not be regarded as a problem: This is precisely what will ensure that the calculated fcc free energy curve lies above the bcc one.

The discussion has so far been concerned with the expression of the partition function, which is the relevant quantity to consider when the phase diagram is calculated with the CVM or the low temperature expansion. Let us now consider the implications of this approach to Monte-Carlo simulations. Thermodynamic quantities derived from averages, such as the average energy, are obviously unaffected by the presence of unstable configurations. For quantities derived from fluctuations, such as the heat capacity, slight modifications are needed. In traditional Monte Carlo simulations, the heat capacity arising from vibrational degrees of freedom is consistently neglected, and any thermodynamic quantity obtained from Monte Carlo can be unambiguously interpreted as the configurational contribution. In the more general setting presented here, there is an overlap between vibrational and configurational fluctuations and the only way to obtain well defined thermodynamic quantities is to fully account for the vibrational fluctuations. Fortunately, there is a straightforward way to do so. The total variance of the energy (or any other quantity) can be exactly expressed as a sum of the variance within each configuration $\vec{\sigma}$ and the variance of

the average energy of each configuration:

$$\left\langle E^2 \right\rangle - \left\langle E \right\rangle^2 = \sum_{\vec{\sigma}} \sum_{v \in \vec{\sigma}} P_{\vec{\sigma}v} E_{\vec{\sigma}v}^2 - \left(\sum_{\vec{\sigma}} \sum_{v \in \vec{\sigma}} P_{\vec{\sigma}v} E_{\vec{\sigma}v} \right)^2$$
$$= \left(\sum_{\vec{\sigma}} P_{\vec{\sigma}} \overline{E}_{\vec{\sigma}}^2 - \left(\sum_{\vec{\sigma}} P_{\vec{\sigma}} \overline{E}_{\vec{\sigma}} \right)^2 \right) + \sum_{\vec{\sigma}} P_{\vec{\sigma}} \left(\sum_{v \in \vec{\sigma}} \frac{P_{\vec{\sigma}v}}{P_{\vec{\sigma}}} \left(E_{\vec{\sigma}v} - \overline{E}_{\vec{\sigma}} \right)^2 \right).$$

where $P_{\vec{\sigma}v}$ is the probability of finding the system in state $\vec{\sigma}$,v while $P_{\vec{\sigma}} = \sum_{v \in \vec{\sigma}} P_{\vec{\sigma}v}$ and $\overline{E}_{\vec{\sigma}} = \sum_{v \in \vec{\sigma}} \frac{P_{\vec{\sigma}v}}{P_{\vec{\sigma}}} E_{\vec{\sigma}v}$. The first term is the usual fluctuation obtained from Monte Carlo. The second term is a correction which takes the form of a simple configuration average of fluctuations within each configuration. The fluctuation of a system constrained to remain in the vicinity of configuration $\vec{\sigma}$ is usually just as simple to determine as its average properties. In the case of energy, the fluctuations within each configuration are simply related the heat capacity of a harmonic solid.

The main objective of this section was to show that there is no fundamental problem associated with unstable states in coarse-graining formalism. While it is true that the free energy of an unstable configuration is not uniquely defined, once a particular way to coarse grain phase space is chosen, the free energy of all configurations can be defined in a consistent fashion. There are admittedly some practical issues to be resolved regarding the practical implementation of coarse graining in the presence of instabilities, but the approach suggested in this section indicates that these difficulties can be overcome.

Appendix B

Structure construction algorithm

B.1 Introduction

This appendix describes an algorithm for constructing a structure from the configuration probabilities describing that structure. This algorithm differs from previous ones in that it can prove the inconstructibility of a potential ground state, instead of simply saying that no structure could be found with unit cell smaller than a given volume. The algorithm gives one of the three following answers:

1. a periodic structure having the required probabilities exists;

2. there exists no structure (periodic or not) having the required probabilities;

3. the algorithm could not decide whether a structure can be found or not.

Note that the last case is rather unlikely. The algorithm can handle a piece of crystal containing hundreds of atoms. If it fails to give a definite answer it means either that the desired periodic structure has hundreds of atoms in its unit cell or that the cluster needed to account for frustration effects contains hundreds of points.

The proposed algorithm initially proceeds as one would do by hand: it attempts to overlap configurations in all possible ways that avoid frustration. What is new, is the way it detects periodicities in the growing structure. Instead of trying all possible unit cells, it simply constructs the structure without assuming any periodic boundary conditions. It then detects in which directions the periodicities are and only tries unit cells having those directions as lattice vectors.

One important limitation of this algorithm: it works only for structures associated with vertices of the configuration polytope obtained from the v-matrix. These so-called "primary vertices" are to be distinguished from "secondary" ones which are obtained after cutting the configuration polytope obtained from the v-matrix with new inequalities obtained using an even bigger maximal cluster.

In this Appendix, we will first motivate the use of our new scheme. We will then present some implementation details that are crucial to make this algorithm practical.

B.2 Basic Principles

Let us consider all supercells whose lattice vectors fit in a cube of $n \times n \times n$ primitive unit cell. There are $(n^3)^3$, or n^9 of them. Symmetry lets us reduce that number, but it will never change the order of the exponent. Hence, to determine the periodicity of a structure using a scheme based on supercell enumeration requires of the order of n^9 operations.

Our algorithm determines periodicities in a structure by testing one-dimensional periodicities. In a cube of $n \times n \times n$ primitive cells there are n^3 possible translation vectors. Among those possible vectors, we select only those which agree with the periodicities that appear to present in the growing structure. (The term "growing structure" refers to the part of the structure that we know at the current stage in the computations.) One can appreciate how small the resulting set of vector is: any vector that maps one atom to an atom of a different type is discarded. Using the few non discarded vectors, we consider all possible triplets of vectors (that is, cells). Then, a more stringent test is applied to the candidate cells. This test essentially amounts to building the structure assuming the periodic boundary conditions dictated by the candidate cell. If such a construction is possible, then a periodic structure has been found. We know (from Section 5.6.8) that the constructed structure must have the right configuration probabilities since we only focus on structures associated with a "primary" vertex.

Note that we avoid testing all possible unit cell by first eliminating lattice vectors which are incompatible with the growing structure. For every discarded lattice vector, we discard (roughly) the n^3 cells having that lattice vector. Note that for small cells, our algorithm can actually be slower than the cell enumeration one (since small cells are typically the first ones to be enumerated). The gains become rapidly appreciable as cell size increases, though. This does not mean that the algorithm is fast: for example, be prepared to wait half an hour[1] for ground states on the fcc lattice with a 10 point octahedron as the maximal cluster.

B.3 Implementation

In order to make the above algorithm fast enough to be practical, one has to take care of the various implementation details which are described in this section. The solutions proposed are not claimed to be optimal. They are simply heuristics that appear to speed up the process in most cases under some plausible assumptions. Note that the implementation choices that were made will never affect the validity of the answer but only the time it takes to find it.

It is nevertheless possible that different implementation choices might give you different structures as an answer. But all the possible structures obtained by changing the heuristics will have the same correlations. This occurs because ground states are often degenerate. In any case, the answer to the question "is there a periodic structure

[1] on a DEC Alpha in 1995 ...

associated with a given set of correlation?" will remain unaffected.

B.3.1 Structure Construction

Exploring the tree of all possible ways to overlap configurations so that they agree on their intersection takes some time. Hence, we need to try the most promising alternatives first. Here are the criterions used to select the "best" way to add a new configuration to the growing structure. This algorithm is recursive: once it has added a new configuration, it calls itself to add the next. If a choice of configuration yields frustration, the current procedure call returns and the parent procedure tries another alternative. In this way, the tree of all possible structure is spanned. The algorithm stops as soon as a periodic structure is found or when the whole tree of possibilities has been spanned.

To add a new configuration to an existing growing structure, we start by finding all possible symmetry operations one can apply to the maximal cluster so that it overlaps at least one of the sites occupied by the growing structure. The "best" way to place the cluster is chosen as following:

1. Maximize the number of sites where the transformed cluster and the growing structure overlap.[2]

2. At equal amount of overlap, take the transformed clusters which are the closest to the origin (this is approximated by the modulus of the lattice translation component of the symmetry operation applied to the cluster).

3. For all clusters located at equal distance from the origin, we now check which configurations can "decorate" a newly added cluster in a way that agrees with the growing structure. We chose clusters which minimize the number of possible configurations. (If the minimum number is actually zero, then this branch yields no constructible structure and another branch has to be tried.)

4. If there are still more than one possible cluster, we arbitrarily take the first one we found.

5. To decorate the newly added cluster, we first try configurations which destroys the minimum amount of periodicity in the growing structure (see below). For each tested configuration we recursively call the construction algorithm (with the growing structure being enlarged by the newly added configuration).

The "amount of periodicity" is evaluated as follows. We first define a candidate lattice vector as being a translation vector such that if one overlaps the growing structure and a translated version of it, similar atoms will align, as illustrated in Figure B-1. Intuitively, if the region where identical atom overlap (gray region in

[2]The trivial case where the transformed cluster lies inside the growing structure is ignored, unless no configuration can be put on that cluster without contradicting the atomic configuration of growing structure.

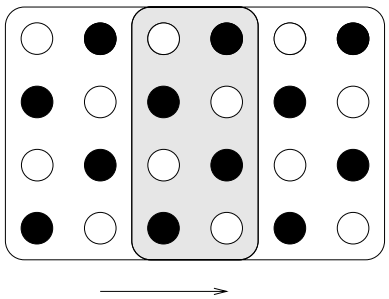


Figure B-1: A Candidate Lattice Vector.

the figure) is larger, the particular periodicity we are considering is more likely to be present in the actual (infinite) structure we are seeking. Hence, a heuristic measure of the "amount of periodicity" can be obtained by summing the areas of the overlapping region (gray region) for all candidate lattice vectors.

B.3.2 Screening Out Candidate Cells

Every time we add a new configuration to the growing structure, we update the status of all translation vectors. For a given growing structure, a particular translation vector can be either

- incompatible with the growing structure (if it maps an atom A to an atom B or vice-versa),
- compatible with the growing structure (*i.e.*, it is a candidate lattice vector) or
- undetermined (the growing structure is not big enough to be able to tell).[3]

Note that it is not sufficient to find three non-coplanar candidate lattice vectors to prove three-dimensional periodicity. A sufficient condition can be expressed in term of the current size and shape of the growing structure. The growing structure must be such that it contains the candidate unit cell plus a "buffer zone" obtained as follows. Apply all possible lattice symmetry operations to the maximal cluster and keep the transformed clusters which overlap the unit cell. Compute the union of all these clusters and keep only sites which have positive coordinates. (Here, we assume that the cell has one of its corners at the origin and that site coordinates are expressed in the basis defined by the candidate unit cell.) This construction is illustrated in Figure B-2 for the 2D case.

The justification of this construction is straightforward. A growing structure of this size (or bigger) can be used to tile space with a periodicity given by the candidate unit cell. The "buffer zone" guaranties that adding new copies of the growing structure will not generate configurations not encountered already. To see this, consider

[3]To increase efficiency, we let translation vectors be undetermined until the region where identical atom overlap is bigger than some threshold (chosen to be the number of sites in the maximum cluster).

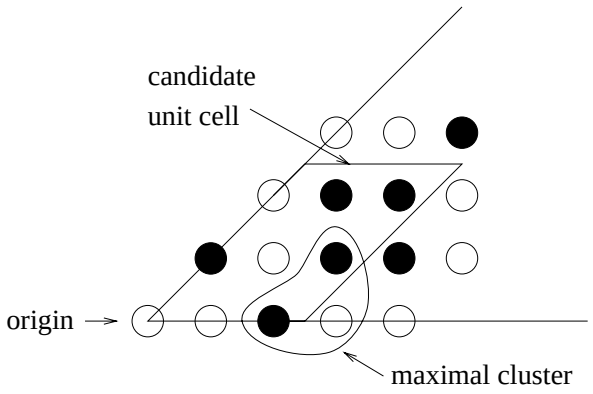


Figure B-2: Sufficient Condition for Two-Dimensional Periodicity.

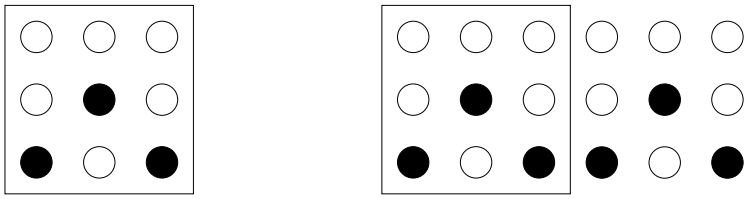


Figure B-3: Problem Arising when no "Buffer Zone" is Used.

Figure B-3, which is an example where no buffer zone is used. The unit cell contains only A-A and A-B bonds while a periodic repetition of it yields B-B bonds.

Since the construction of a "buffer zone" is rather computer intensive, we need a way to avoid this test as much as possible. Starting with the list of candidate lattice vectors we go through a sequence of increasingly more stringent (and more computer intensive) tests to reduce the number of candidate unit cells. If a test fails, the subsequent ones are not applied.

1. We take all triplets a, b, c of non coplanar vectors.

2. We check if $a + b$, $b + c$, $a + c$ and $a + b + c$ are also candidate vectors.

3. We check if the resulting cell has a volume compatible with the concentration of each atom type.

4. We then use the length of the largest line segment that can fit in the unit cell to quickly see if it would fit in the growing structure.[4]

5. Finally, the full "buffer zone" test is applied.

Parts of this screening process are really heuristic. Again, this is not a problem, because either of the following can occur in the worst case.

[4]The width of the buffer zone is added to the length of the line segment. This width is approximated by the radius of gyration of the maximal cluster.

- Some heuristic tests might be too stringent at some point and screen out perfectly valid unit cells. However, all tests are such that, if three-dimensional periodicity exists, they will succeed for a big enough growing structure. They may fail at an early time, but will always succeed eventually, if a not too big periodic structure exists. The only drawback is a loss of speed.

- Some heuristic tests might be too lax, it which case the full, exact, "buffer zone" test is applied too often. Again, rigor is not lost, only speed.

B.4 Conclusion

The proposed algorithm enables us to determine whether a potential ground state in constructible or not. It will fail to give a definite answer only if a ground state's unit cell is very large or if frustration effects have a very long range.

It is important to keep in mind that this algorithm is limited to the problem of constructing ground state structures. For arbitrary structures, the cell enumeration algorithm is still the method of choice. Also, if the unit cell is known to be small or if we want to be sure to obtain the smallest possible unit cell, the cell enumeration algorithm should again be chosen.

For large unit cells (*e.g.*, 30 atoms per unit cell) our algorithm is faster than if one had to enumerate all unit cells. However, this can still be slow from a practical point of view (maybe a few hours of computer time). It is probably possible to refine the heuristics used to speed up the algorithm, but the present implementation seems a reasonable compromise between complexity and speed.

As a by-product of this constructing code, a large library of functions related to lattices and symmetry operations on them has been developed. It was made sufficiently general to be easily used in other contexts.

Bibliography

[1] I. A. Abrikosov, A. V. Ruban, D. Ya Kats, and Yu. H. Vekilov. Electronic structure, thermodynamic and thermal properties of Ni–Al disordered alloys from LMTO–CPA–DFT calculations. *Journal of Physics: Condensed Matter*, 5:1271–90, 1993.

[2] G. J. Ackland. Vibrational entropy of ordered and disordered alloys. In G. Stocks and P. Turchi, editors, *Alloy Modelling and Design*, pages 149–153, Pittsburgh, PA, 1994. The Minerals, Metals and Materials Society.

[3] G. J. Ackland, M. C. Warren, and S. J. Clark. Practical methods in *ab-initio* lattice dynamics. *J. Phys.: Condens. Matter*, 9:7861–7872, 1997.

[4] I. V. Alexandrov and R. Z. Valiev. Computer simulation of X-ray diffraction patterns of nanocrystalline materials. *Phil. Mag. B*, 73(6):861–872, 1996.

[5] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids*. Clarendon Press, Oxford, 1987.

[6] S. M. Allen and J. W. Cahn. Ground state structures in ordered binary alloys with second neighbor interactions. *Acta Metall.*, 20:423–433, 1972.

[7] C.-O. Almbladh and A. C. Pedroza. Density-functional exchange-correlation potentials and orbital eigenvalues for light atoms. *Phys. Rev. A*, 29(5):2322–2330, 1984.

[8] J. D. Althoff, D. Morgan, D. de Fontaine, M. Asta, S. M. Foiles, and D. D. Johnson. Vibrational spectra in ordered and disordered ni$_3$al. *Phys. Rev. B*, 56:R5705, 1997.

[9] J. D. Althoff, D. Morgan, D. de Fontaine, M. D. Asta, S. M. Foiles, and D. D. Johnson. Embedded atom method calculations of vibrational thermodynamics properties of ordered and disordered Ni_3Al. *Comput. Mater. Sci.*, 10:411–415, 1998.

[10] L. Anthony, L. J. Nagel, J. K. Okamoto, and B. Fultz. Magnitude and origin of the difference in vibrational entropy between ordered and disordered Fe_3Al. *Phys. Rev. Lett.*, 73(22):3034–3037, 1994.

[11] L. Anthony, J. K. Okamoto, and B. Fultz. Vibrational entropy of ordered and disordered Ni_3Al. *Phys. Rev. Lett.*, 70(8):1128–1130, 1993.

[12] Neil W. Ashcroft and N. David Mermin. *Solid State Physics*. Harcourt Brace College Publisher, New York, 1976.

[13] M. Asta, R. McCormack, and D. de Fontaine. Theoretical study of alloy stability in the cd-mg system. *Phys. Rev. B*, 48(2):748–766, 1993.

[14] H. Bakker. Numerical calculations of the change of the vibrational entropy due to disordering of a one- and two-dimensional ordered binary alloy. *Phil. Mag. A*, 45(1):213–219, 1982.

[15] H. Bakker. The vibrational entropy change associated with the disordering of a two-dimensional binary alloy using a model of nearest- and next-nearest-neighbour intereactions. *Phys. Stat. Sol. (b)*, 109:211–216, 1982.

[16] H. Bakker and C. Tuijn. Extension of the ising model for calculations on order-disorder phenomena in alloys by allowing for atomic vibrations. *J. Phys. C: Solid State Phys.*, 19:5585–5589, 1986.

[17] S. Baroni, P. Giannozzi, and A. Testa. Green's-function approach to linear response in solids. *Phys. Rev. Lett.*, 58(18):1861–1864, 1987.

[18] J. D. Becker and J. M. Sanchez. First principles phase stability study of the Ru–Nb–Zr system. *Materials Science & Engineering A (Structural Materials: Properties, Microstructure and Processing)*, A170:161–7, 1993.

[19] K. Binder and D. W. Heermann. *Monte Carlo Simulation in Statistical Physics*. Springer-Verlag, New York, 1988.

[20] P. D. Bogdanoff and B. Fultz. Vibrational entropies of alloying and compound formation: experimental trends. *Phil. Mag. B*, 79(5):753–765, 1999.

[21] P. D. Bogdanoff, B. Fultz, and S. Rosenkranz. Vibrational entropy of $L1_2$ Cu_3Au measured by inelastic scattering. *Phys. Rev. B*, 60(6):3976–3981, 1999.

[22] D. Le Bolloc'h, R. Caudron, and A. Finel. Experimental and theoretical study of the temperature and concentration dependence of the short-range order in pt-v alloys. *Phys. Rev. B*, 57(5):2801–2811, 1998.

[23] C. Booth and J. S. Rowlinson. The influence of lattice vibrations on the order-disorder transitions of alloys. *Trans. Faraday Soc.*, 51:463–467, 1955.

[24] M. Born and K. Huang. *Dynamical Theory of Crystal Lattices*. Ocford University Press, London, 1956.

[25] L. L. Boyer and M. J. Mehl. A self consistent atomic deformation model for total energy calculations: application to ferroelectrics. *Ferroelectrics*, 150:13–24, 1993.

[26] F. Cardelli, F. Cleri, G. Mazzone, A. Montone, and V. Rosato. Experimental and theoretical investigation of the order-disorder transformation in Ni_3Al. *J. of Mater. Res.*, 8(10):2504–2509, 1993.

[27] G. Ceder. *Alloy theory and its application to long period superstructures ordering in metalic alloys and high temperature superconductors.* PhD thesis, University of California at Berkeley, 1991.

[28] G. Ceder. A derivation of the ising model for the computation of phase diagrams. *Comput. Mater. Sci.*, 1:144–150, 1993.

[29] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45(7):566–569, 1980.

[30] D. J. Chadi and M. Cohen. Special points in the brillouin zone. *Phys. Rev. B*, 8(12):5747–5753, 1973.

[31] V. Chvatal. *Linear programming.* W.H. Freeman, San Francisco, 1983.

[32] F. Cleri and V. Rosato. Lattice dynamics of ordered and disordered Cu_3Au with a tight-binding potential model. *Phil. Mag. Lett.*, 67(6):369–378, 1993.

[33] R. E. Cohen. Origin of ferroelectricity in perovskite oxides. *Nature*, 358:136–138, July 1992.

[34] C. Colinet, J. Eymery, A. Pasturel, A. T. Paxton, et al. A first–principles phase stability study on the Au–Ni system. *Journal of Physics: Condensed Matter*, 6:L47–52, 1994.

[35] A. Dal Corso, A. Pasquarello, and A. Baldereschi. Density-functional perturbation theory for lattice dynamics with ultrasoft pseudopotentials. *Phys. Rev. B*, 56(18):R11369–R11372, 1997.

[36] P. J. Craievich and J. M. Sanchez. Vibrational free energy in the ni-cr system. *Comput. Mater. Sci.*, 8:92–99, 1997.

[37] P. J. Craievich, J.M. Sanchez, R. E. Watson, and M. Weinert. Structural instabilities of excited phases. *Phys. Rev. B*, 55(2):787–797, 1997.

[38] M. S. Daw and M. I. Baskes. Embedded–atom method: derivation and application to impurities, surfaces and other defects in metals. *Physical Review B*, 29:6443, 1984.

[39] D. de Fontaine. Cluster approach to order-disorder transformation in alloys. *Solid State Phys.*, 47:33–176, 1994.

[40] S. de Gironcoli. Lattice dynamics of metals from density–functional perturbation theory. *Phys. Rev. B*, 51:6773–6776, 1995.

[41] F. Ducastelle. *Order and Phase Stability in Alloys*. Elsevier Science, New York, 1991.

[42] F. J. Dyson. The dynamics of a disordered linear chain. *Physical Review*, 92:1331–1338, 1953.

[43] M. Ernzerhof, J. P. Perdew, and K. Burke. Density functionals: Where do they come from, why do they work? *Top. Curr. Chem.*, 180:1–29, 1996.

[44] S. Fahy, X. W. Wang, and S. G. Louie. Pair-correlation function and single-particle occupation numbers in diamond and silicon. *Phys. Rev. Lett.*, 65(12):1478–1481, 1990.

[45] C. Filippi, D. J. Singh, and C. J. Umrigar. All-electron local-density and generalized-gradient calculations of the structural properties of semiconductors. *Phys. Rev. B*, 50(20):14947–14951, 1994.

[46] D. De Fontaine, J. Althoff, D. Morgan, M. Asta, S. Foiles, and D. Johnson A. Quong. Approximations for vibrational thermodynamics of disordered alloys: effective supercells and the quasiharmonic method. In E. Ma, M. Atzmon, P. Bellon, and R. Trivedi, editors, *Phase Transformations and Systems Driven Far From Equilibrium*, pages 175–82, Warrendale, PA, 1998. Mater. Res. Soc.

[47] H. N. Frase, L. J. Nagel, J. L. Robertson, and B. Fultz. Vibrational density of states of nanocrystalline ni$_3$fe. *Phil. Mag. B*, 75(3):335–347, 1997.

[48] M. Fuchs, M. Bockstedte, E. Pehlke, and M. Scheffler. Pseudopotential study of binding properties of solids within generalized gradient approximation: The role of core-valence exchange correlation. *Phys. Rev. B*, 57(4):2134–2145, 1998.

[49] B. Fultz, L. Anthony, L. J. Nagel, J. K. Okamoto, R. M. Nicklow, and S. Spooner. Phonon density of states and vibrational entropies of ordered and disordered Ni_3Al. *Phys. Rev. B*, 52(5):3315–3321, 1995.

[50] B. Fultz, L. Anthony, J. L. Robertson, R. M. Nicklow, S. Spooner, and M. Mostoller. Phonon modes and vibrational entropy of mixing in Fe-Cr. *Phys. Rev. B*, 52(5):3280–3285, 1995.

[51] B. Fultz, J. L. Robertson, T. A. Stephens, L. J. Nagel, and S. Spooner. Phonon density of states of nanocrystalline fe prepared by high-energy ball milling. *J. Appl. Phys.*, 79(11):8319–8322, 1996.

[52] B. Fultz, T. A. Stephens, W. Sturhahn, T. S. Toellner, and E. E. Alp. Local chemical environment and the phonon partial densities of states of ^{57}Fe in $^{57}Fe_3Al$. *Phys. Rev. Lett.*, 80(15):3304–3307, 1998.

[53] G. D. Garbulsky and G. Ceder. Effect of lattice vibrations on the ordering tendencies in substitutional binary alloys. *Phys. Rev. B*, 49:6327–6330, 1994.

[54] G. D. Garbulsky and G. Ceder. Contribution of the vibrational free energy to phases stability in substitutional alloys: methods and trends. *Phys. Rev. B*, 53:8993–9001, 1996.

[55] Gerardo D. Garbulsky. *Ground-state structures and vibrational free energy in first-principles models of substitutional-alloy thermodynamics*. PhD thesis, Massachusetts Institute of Technology, 1996.

[56] P. Giannozzi, S. de Gironcoli, P. Pavone, and S. Baroni. Ab initio calculation of phonon dispersions in semiconductors. *Phys. Rev. B*, 43:7231, 1991.

[57] X. Gonze, D. C. Allan, and M. P. Teter. Dielectric tensor, effective charges, and phonons in α-quartz by variational density-functional perturbation theory. *Phys. Rev. Lett.*, 68:3603, 1992.

[58] G. Grimvall. *Thermophysical Properties of Materials*. Selected Topics in Solid State Physics. North-Holland, 1986.

[59] G. Grimvall and I. Ebbsjo. Polymorphism in metals. I. vibrational free energy. *Physica Scripta*, 12:168, 1975.

[60] O. V. Gritsenko, R. van Leeuwen, and E. J. Baerends. Molecular kohn-sham exchange-correlation potential from the correlated ab initio electron density. *Phys. Rev. A*, 52(3):1870–1874, 1995.

[61] S. R. Harris, D. H. Pearson, C. M. Garland, and B. Fultz. Chemically disordered ni$_3$al synthesized by high vacuum evaporation. *J. Mater. Res.*, 6(10):2019–2021, 1991.

[62] W. A. Harrison. *Electronic Structure and the Properties of Solids*. Dover Publications, New York, 1989.

[63] K. C. Hass, L. C. Davis, and A. Zunger. Electronic structure of random al$_{0.5}$ga$_{0.5}$as alloys: test of the 'special-quasirandom-structure' description. *Phys. Rev. B*, 42(6):3757–3760, 1990.

[64] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864, 1964.

[65] R. Q. Hood, M. Y. Chou, A. J. Williamson, G. Rajagopal, R.J. Needs, and W. M. C. Foulkes. Quantum monte carlo investigation of exchange and correlation in silicon. *Phys. Rev. Lett.*, 78(17):3350–3353, 1997.

[66] J. S. C. Jang and C. C. Koch. Amorphization and disordering of the ni$_3$al ordered intermetallic by mechanical milling. *J. Mater. Res.*, 5(3):498–509, 1990.

[67] R. O. Jones and O. Gunnarsson. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, 61(3):689–746, 1989.

[68] A. Khein, D. J. Singh, and C. J. Umrigar. All-electron study of gradient corrections to the local-density functional in metallic systems. *Phys. Rev. B*, 51(7):4105–4109, 1995.

[69] R. Kikuchi. A theory of cooperative phenomena. *Phys. Rev.*, 81:988, 1951.

[70] A. F. Kohan, P. D. Tepesch, G. Ceder, and C. Wolverton. Computation of alloy phase diagrams at low temperatures. *Comp. Mat. Sci.*, 9:389–396, 1998.

[71] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133, 1965.

[72] W. Kohn and P. Vashishta. General density functional theory. In Lundqvist and March [80], pages 79–147.

[73] X. J. Kong, C. T. Chan, K. M. Ho, and Y.Y. Ye. Cohesive properties of crystalline solids by the generalized gradient approximation. *Phys. Rev. B.*, 42(15):9357–9364, 1990.

[74] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comp. Mat. Sci.*, 6:15, 1996.

[75] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169, 1996.

[76] N. D. Lang. Density functional approach to the electronic structure of metal surfaces and metal-adsorbate systems. In Lundqvist and March [80], pages 309–376.

[77] I.-H. Lee and R. M. Martin. Application of the generalized-gradient approximation to atoms, clusters, and solids. *Phys. Rev. B*, 56(12):7197–7205, 1997.

[78] R. LeSar, R. Najafadabi, and D. J. Srolovitz. Finite–temperature defect properties from free energy minimization. *Physical Review Letters*, 63:624, 1989.

[79] K. Lu, H. Y. Zhang, Y. Zhong, and H. J. Fecht. Grain size dependence of mechanical properties in nanocrystalline selenium. *J. Mater. Res.*, 12(4):923–930, 1997.

[80] S. Lundqvist and N. H. March, editors. *Theory of the Inhomogeneous Electron Gas*, New York, 1983. Plenum Press.

[81] A. A. Maradudin, E. W. Montroll, and G. H. Weiss. *Theory of Lattice Dynamics in the Harmonic Approximation, Second Edition.* Academic Press, New York, 1971.

[82] J. A. D. Matthew, R. E. Jones, and V. M. Dwyer. The effect of short-range order on the vibrational entropy of one-dimensional chains. *J. Phys. F: Met. Phys.*, 13:581–585, 1983.

[83] S. K. Mishra and G. Ceder. Structural stability of lithium manganese oxides. *Phys. Rev. B*, 59(9):in press, 1999.

[84] T. Mohri. Theoretical study on the phase stability of a III–V semiconductor alloy. *Progress of Theoretical Physics Supplement*, 115:147–64, 1994.

[85] T. Mohri, S. Takizawa, and K. Terakura. First-principles calculation of heats of formation for Au–Cu, Au–Pd and Au–Ag alloys with thermal vibration effects. *Journal of Physics: Condensed Matter*, 5:1473–80, 1993.

[86] H. J. Monkhorst and J. Pack. Special points for brillouin-zone integration. *Phys. Rev. B*, 13(12):5188–5192, 1976.

[87] G. Moraisis and F. Gautier. Excess vibrational entropy and local environment effects. *J. Phys. F: Metal Phys.*, 7(8):1421–1433, 1977.

[88] D. Morgan. Private Communication.

[89] D. Morgan. *Computational Studies of Alloy Phase Stabilities.* PhD thesis, University of California at Berkeley, 1998.

[90] D. Morgan, J. D. Althoff, and D. de Fontaine. Local environment effects in the vibrational properties of disordered alloys: an embedded-atom method study of ni$_3$al and cu$_3$au. *J. of Phase Equilibria*, 19(6):559–567, 1998.

[91] D. Morgan and A. van de Walle and. Vibrational thermodynamics: coupling of chemical order and size effects. *Modelling Simul. Mater Sci Eng.*, 8:1–5, 2000.

[92] V. L. Moruzzi, J. F. Janak, and K. Schwarz. Calculated thermal properties of metals. *Phys. Rev. B*, 37:790–799, 1988.

[93] G. D. Mukherjee, C. Bansal, and A. Chatterjee. Thermal expansion study of ordered and disordered Fe_3Al: An effective approach for the determination of vibrational entropy. *Phys. Rev. Lett.*, 76:1876, 1996.

[94] G. D. Mukherjee, C. Bansal, and A. Chatterjee. Vibrational and magnetic contribution to entropy of an $L1_2$ ordered and a partially ordered Ni_3Mn alloy: Results from thermal expansion measurements and model calculations. *International J. of Mod. Phys. B*, 12(22):2233, 1998.

[95] L. J. Nagel, L. Anthony, and B. Fultz. Differences in vibrational entropy of disordered and ordered Cu_3Al. *Phil. Mag. Lett.*, 72(6):421–427, 1995.

[96] L. J. Nagel, L. Anthony, J. K. Okamoto, and B. Fultz. An experimental study of the difference in vibrational entropy between ordered and disordered Fe_3Al. *J. Phase Equilibria*, 18(6):551–555, 1997.

[97] L. J. Nagel, B. Fultz, and J. L. Robertson. Vibrational entropies of phases of Co_3V measured by inelastic neutron scattering and cryogenic calorimetry. *Phil. Mag. B*, 75(5):681–699, 1997.

[98] L. J. Nagel, B. Fultz, J. L. Robertson, and S. Spooner. Vibrational entropy and microstructural effects on the thermodynamics of partially disordered and ordered Ni_3V. *Phys. Rev. B*, 55(5):2903–2911, 1996.

[99] R. Najafabadi, H. Y. Wang, D. J. Srolovitz, and R. LeSar. A new method for the simulation of alloys: application to interfacial segregation. *Acta Metallurgica et Materialia*, 39:3071–82, 1991.

[100] K. Nakamura and T. Mohri. Non–empirical calculation of thermal vibration effects on the phase stability of InP–InSb alloy. *Modelling and Simulation in Materials Science and Engineering*, 1:143–50, 1993.

[101] Chang-Seok Oh, T. Mohri, and Dong Nyung Lee. Phenomenological calculation of the $L1_0$–disorder phase equilibria in the Co–Pt system. *Materials Transactions JIM*, 35:445–50, 1994.

[102] Joint Committee on Powder Diffraction Standards. *Powder diffraction file: inorganic volume*. JCPDS International Centre for Diffraction Data, Swarthmore, PA, 1970.

[103] G. Ortiz and P. Ballone. Pseudopotential for non-local-density functionals. *Phys. Rev. B*, 43(8):6376–6387, 1991.

[104] V. Ozoliņš. *Structural and Vibrational Properties of Transition Metal Systems from ab initio Electronic-Structure Calculations*. PhD thesis, Royal Institute of Technology, Stockholm, Sweden, 1996.

[105] V. Ozoliņš and M. Körling. Full-potential calculations using the generalized gradient approximation: Structural properties of transition metals. *Phys. Rev. B*, 48(24):18304–18307, 1993.

[106] V. Ozoliņš, C. Wolverton, and A. Zunger. First-principles theory of vibrational effects on the phase stability of cu-au compounds and alloys. *Phys. Rev. B*, 58(10):R5897–R5900, 1998.

[107] W. B. Pearson, P. Villars, and L.D. Calvert. *Pearson's handbook of crystallographic data for intermetallic phases*. American Society for Metals, Metals Park, OH, 1985.

[108] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. Atoms, molecules, solids and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys Rev. B*, 46(11):6671–6687, 1992.

[109] J. P. Perdew and A. Zunger. Self-interaction correction to density-functinal approximation for many-electron systems. *Phys. Rev. B*, 23(10):5048–5079, 1981.

[110] K. Persson, M. Ekman, and G. Grimvall. Dynamical and thermodynamical instabilities in the disordered Re_xW_{1-x} system. *Phys. Rev. B*, 60(14):9999–10007, 1999.

[111] W. Petry, A. Heiming, J. Trampenau, M. Alba, C. Herzig, H. R. Schober, and G. Vogl. Phonon dispersion of the bcc phase of group-iv metals. i. bcc titanium. *Phys. Rev. B*, 43(13):10933–10947, 1991.

[112] D. G. Pettifor. The tight binding approximation: Concepts and predictions. In D.G. Pettifor and A. H. Cottrell, editors, *Electron theory in alloy design*, page 81, London, 1992. The Institute of Materials.

[113] J. C. Phillips and L. Kleinman. New method for calculating wave functions in crystals and molecules. *Phys. Rev.*, 116:287–294, 1959.

[114] A. A. Quong and B. Klein. Self-consistent-screening calculation of interatomic force constants and phonon dispersion curves from first principles: Application to aluminum. *Phys. Rev. B*, 46:10734, 1992.

[115] R. Ravelo, J. Aguilar, M. Baskes, J.E. Angelo, B. Fultz, and B. L. Holian. Free energy and vibrational entropy difference between ordered and disordered ni_3al. *Phys. Rev. B*, 57:862, 1998.

[116] J. D. Rittner, S. M. Foiles, and D. N. Seidman. Simulation of surface segregation free energies. *Physical Review B*, 50:12004–14, 1994.

[117] J. L. Robertson, B. Fultz, and H. N. Frase. Phonon contributions to the entropies of $hP24$ and fcc Co_3V. *Phys. Rev. B*, 60(13):9329–9334, 1999.

[118] J. M. Sanchez and J. D. Becker. The role of the cluster variation method in the first principles calculation of phase diagrams. *Prog. Theor. Phys. Supp.*, 115:45, 1994.

[119] J. M. Sanchez, F. Ducastelle, and D. Gratias. Generalized cluster description of multicomponent systems. *Physica*, 128A:334, 1984.

[120] J. M. Sanchez, J. P. Stark, and V. L. Moruzzi. First-principles calculation of the ag-cu phase diagram. *Phys. Rev. B*, 44(11):5411–5418, 1991.

[121] L. Shaojun, D. Suqing, and M. Benkun. First-principles calculation of vibrational entropy for fe-al compounds. *Phys. Rev. B*, 58(15):9705–9709, 1998.

[122] A. Silverman, A. Zunger, A. Kalish, and J. Adler. Atomic–scale structure of disordered $Ga_{1-x}In_xP$ alloys. *Physical Review B*, 51:10795–816, 1995.

[123] M. H. Sluiter, M. Weinert, and Y. Kawazoe. Force constants for substitutional alloys. *Phys. Rev. B*, 59(6):4100–4111, 1999.

[124] F. Solal, R. Caudron, F. Ducastelle, A. Finel, and A. Loiseau. Long-range and short-range order in pd_3v: Breakdown of the mean- field theory. *Phys. Rev. Lett.*, 58(21):2245–2248, 1987.

[125] R. Stadler, W. Wolf, R. Podloucky, G. Kresse, J. Furthmüller, and J. Hafner. Ab initio calculations of the cohesive, elastic, and dynamical properties of $cosi_2$. *Phys. Rev. B*, 54(3):1729–1733, 1996.

[126] C. Stassis, F. X. Kayser, C.-K. Loong, and D. Arch. Lattice dynamics of ni_3al. *Phys. Rev. B*, 24(6):3048–3053, 1981.

[127] G. Sugiyama, G. Zerah, and B. J. Alder. Ground-state properties of metallic lithium. *Physica A*, 156:144–168, 1989.

[128] A. P. Sutton. Temperature–dependent interatomic forces. *Philosophical Magazine A*, 60:147–59, 1989.

[129] P. D. Tepesch, A. F. Kohan, G. D. Garbulsky, and G. Ceder *et al.* A model to compute phase diagrams in oxides with empirical or first-principles energy methods and application to the solubility limits in the cao–mgo system. *J. Am. Ceram.*, 49:2033–2040, 1996.

[130] Y. S. Touloukian. *Thermophysical Properties of Matter, TPRC Data Series.* IFI/Plenum, New York, 1970-1979.

[131] J. Trampenau, K. Bauszus, W. Petry, and U. Herr. Vibrational behavior of nanocrystalline ni. *Nanostruct. Mater.*, 6:551–554, 1995.

[132] W. T. Tseng and J. P. Stark. First-principle calculations of solid solubility of titanium in aluminum alloys. *Phil. Mag. B*, 70(4):919–926, 1994.

[133] C. Tuijn and H. Bakker. The influence of the vibrational entropy on the specific heat of the order-disorder transition in alloys. *Phys. Stat. Sol. (b)*, 155:107–112, 1989.

[134] A. van de Walle. Relating fractal image compression to transform methods. Master's thesis, University of Waterloo, Canada, 1995.

[135] A. van de Walle and G. Ceder. Correcting overbinding in lda calculations. *Phys. Rev. B*, 59:14992–15001, 1999.

[136] A. van de Walle, G. Ceder, and U. V. Waghmare. First-principles computation of the vibrational entropy of ordered and disordered ni_3al. *Phys. Rev. Lett.*, 80(22):4911–4914, 1998.

[137] A. van de Walle and G.Ceder. First-principles computation of the vibrational entropy of ordered and disordered Pd_3V. *Phys. Rev. B*, 61:5972–5978, 2000.

[138] R. van Leeuwen, O. V. Gritsenko, and E. J. Baerends. Analysis and modeling of atomic and molecular kohn-sham potentials. *Top. Curr. Chem.*, 180:107–167, 1996.

[139] D. Vanderbilt. Soft self–consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, 41:7892–7895, 1990.

[140] A. F. Voter and S. P. Chen. Characterization of defects in materials. In R. W. Siegel, J. R. Weertman, and R. Sinclair, editors, *Characterization of Defects in materials*, page 175, Pittsburg, 1988. Materials Research Society.

[141] A. A. H. J. Waegmaekers and H. Bakker. The change of the vibration entropy due to the disordering of an ordered simple–cubic alloy. In *Materials Research Society Symposium Proceedings*, volume 21, page 343, 1984.

[142] U. V. Waghmare. *Ab Initio Statistical Mechanics of Structural Phase Transitions (Lattice Dynamics, Piezoelectric, Ferroelectric)*. PhD thesis, Yale University, 1996.

[143] S. Wei and M. Y. Chou. Ab initio calculation of force constants and full phonon dispersions. *Phys. Rev. Lett.*, 69:2799, 1992.

[144] P. J. Wojtowicz and J. G. Kirkwood. Contribution of lattice vibrations to the order-disorder transformation in alloys. *J. Chem. Phys.*, 33(5):1299–1310, 1960.

[145] C. Wolverton and A. Zunger. First-principles theory of short-range order, electronic excitations, and spin polarization in ni-v and pd-v alloys. *Phys. Rev. B*, 52(12):8813–8828, 1995.

[146] L. Zhao, R. Najafabadi, and D. J. Srolovitz. Finite temperature vacancy formation thermodynamics: local harmonic and quasiharmonic studies. *Modelling and Simulation in Materials Science and Engineering*, 1:539–51, 1993.

[147] W. Zhong and D. Vanderbilt. Competing structural instabilities in cubic perovskites. *Phys. Rev. Lett.*, 74(13):2587–2590, 1995.

[148] W. Zhong, D. Vanderbilt, and K. M. Rabe. Phase transition in batio$_3$ from first principles. *Phys. Rev. Lett.*, 73(13):1861–1864, 1994.

[149] V. I. Zubov. Surface properties of solids and size effects in nanophases. *Nanostruct. Mater.*, 3:189–193, 1993.

[150] A. Zunger. First principles statistical mechanics of semiconductor alloys and intermetallic compounds. In P. E. Turchi and A. Gonis, editors, *NATO ASI on Statics and Dynamics of Alloy Phase Transformation*, volume 319, pages 361–419, New York, 1994. Plenum Press.

[151] Alex Zunger, S.-H. Wei, L.G. Ferreira, and James E. Bernard. Special quasirandom structures. *Phys. Rev. Lett.*, 65(3):353–356, 1990.